Investigations into an Archaeal RNA Polymerase: Structure to Function Analysis

A dissertation by

Maria-Elena Mogni

Submitted to the University of Oxford in partial satisfaction of the requirements

for the degree of DOCTOR OF PHILOSOPHY

Major-Subject: Chromosome and Developmental Biology

Supervision: Prof. Stephen Bell

Wadham College

October 2012

Abstract

Investigations into an Archaeal RNA Polymerase: Structure to Function Analysis

Maria Elena Mogni - Wadham College - University of Oxford

Submitted for the degree of Doctor of Philosophy in Michaelmas Term 2012

The archaeal RNA polymerase (RNAP) is similar to the eukaryotic RNAP-II in terms of subunit composition and overall protein structure. Despite its similarity, a new archaeal-specific Rpo13 subunit has been identified. Rpo13 occupies a position in the enzyme which, in RNAP-II, is filled by the eukaryotic-specific Rpb5 jaw domain. It has therefore been proposed to contact DNA, where the positively charged C-terminal tail might mediate protein-DNA interactions. Furthermore, analysis of archaeal genomes has identified a homologue of the eukaryotic RNAP-III-specific RPC34 subunit. RPC34 may associate with the single archaeal RNAP, modulating the specificity of the archaeal RNAP and re-directing it to a subset of genes such as non-coding genes, in analogy to the RNAP-III/RPC34 eukaryotic system involved in the transcription of 5S rRNA, tRNAs and other small RNAs. More importantly, the association of RPC34 with the single archaeal RNAP would define an archaeal enzyme which acts as a precursor of the eukaryotic RNAP-III. Electrophoretic mobility shift assay (EMSA) analysis of purified Rpo13 protein by recombinant means subsequently incubated with a double-stranded (ds)DNA sequence reveals the formation of protein-DNA complexes, where Rpo13's binding to DNA is non-sequence specific but discriminatory to dsDNA, as no binding is observed in the presence of single-stranded (ss)DNA. Also, it is found that the

major determinant of DNA binding is the Rpo13's positively charged C-terminal tail, since DNA binding is abolished with a Rpo13 mutant deficient in this tail. Furthermore, neither major groove nor minor groove interacting compounds have a major impact on Rpo13's binding to DNA, suggesting that Rpo13 may associate with the negatively charged DNA phosphate backbone. Moreover, *in vitro* transcription assays indicate that a transcription product is observed upon RNAP incubation with a bubble DNA oligo shown to make Rpo13 contacts in the RNAP-DNA crystal structure. In addition, while a GST-pulldown experiment suggests the existence of an interaction between the archaeal RNAP and RCP34 *in vitro*, co-immunoprecipitation assays argue against the existence of such interaction from an *in vivo* point of view. Finally, a chromatin immunoprecipitation (ChIP)-sequencing approach to analyse Rpo13's genomic distribution versus the one of the bulk RNAP was undertaken. While the gel filtration elution profile analysis of Rpo13 in the *S. acidocaldarius* cell extract versus the one of recombinant Rpo13 suggests that there is a free Rpo13 pool in the cell extract, indicating that Rpo13 may be acting as a transiently-associated RNAP subunit displaying a factor-like function, the ChIP-sequencing approach reveals that Rpo13 is a *bona fide* RNAP subunit since it co-localises from a genomic point of view with the bulk RNAP.

Declaration

The work described in this thesis was carried out between October 2009 and October 2012, in the laboratory of Prof. Stephen Bell at the Sir William Dunn School of Pathology of the University of Oxford. All the work described within this thesis is entirely my own unless stated. This work has not been submitted previously for any other degree at the University of Oxford or any other university. Publications stemming from this work are detailed prior to the document contents.

Maria Elena Mogni

October 2012

List of Publications

1. Mogni ME, Costa A, Ioannou C, Bell SD. The glutamate switch is present in all seven clades of AAA+ protein. *Biochemistry*, 48 (2009), 8774–5.

2. Wojtas M, Peralta B, Ondiviela M, Mogni M, Bell SD, Abrescia NG. Archaeal RNA polymerase: the influence of the protruding stalk in crystal packing and preliminary biophysical analysis of the Rpo13 subunit. *Biochem Soc. Trans.*, 39 (2011), 25–30.

3. Wojtas MN, Mogni M, Millet O, Bell SD, Abrescia NG. Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA. *Nucleic Acids Res.* (2012), [Epub ahead of print]

Acknowledgments

First of all, I would like to thank Prof. Stephen Bell for his supervision of my D.Phil project these three years. Thanks to him, I have learned a lot of new science techniques such as molecular biology techniques, protein purification and protein biochemistry assays, and developed my scientific thinking. Thanks also go to the rest of the laboratory, in no particular order, to Dr. Tamzin Gristwood, Dr. Nelly Dubarry, Dr. Ben Hogson, Dr. Rachel Samson, Dr. Li Phing, Dr. Helen Farr, Dr. Germe and pro-to-doctor Thomas Beattie. Particularly, I would like to thank Dr. Ben Hogson for all the media preparation and for teaching me how to affinity purify antibodies. I would like to also thank Dr. Rachel Samson for teaching me how to prepare RNA samples from a *S. acidocaldarius* cell extract and for guiding me through the ChIP-sequencing procedure. Moreover, I would like to thank Thomas Beattie for helping me in trouble shooting all the problems related to the malfunctioning of research instruments and devices. All together, their patience and support has been of immense help for me to carry on and to keep my motivation high. In addition, I would like to thank our collaborators Nicola Abrescia and Magda Wojtas for all of their useful discussions, as well as for providing the RNA polymerase sample and the TBP-TFB fusion protein. Finally, I want to thank my Mother and Father for supporting me during my studies, to whom I dedicate this thesis. Special thanks also go to the Wellcome Trust for their generous funding for this D.Phil project.

Contents

1 Introduction **1**

1.1 The Archaeal Domain of Life 1

1.2 Transcription as a Cellular Process 4

1.3 Eukaryotic Transcription 6

1.3.1 RNAP-II's Overall Structure 6

1.3.2 RNAP-II's Catalytic Elements 8

1.3.3 RNAP-II's Rpb4/Rpb7 Complex 10

1.3.4 Catalysis and Transcription Elongation by RNAP-II 12

1.3.5 Transcription Initiation by RNAP-II 13

1.3.6 Other Eukaryotic Polymerases 21

1.4 Archaeal Transcription 23

1.4.1 Archaeal RNAP's Overall Structure 24

1.4.2 Archaeal Transcription Factors 29

2 Experimental Aims **32**

3 Materials and Methods **36**

3.1 Cloning 36

3.2 Recombinant Protein Purification 37

3.3 Antibody Raising against Recombinant Rpo13 and RPC34 42

3.4 EMSA Analysis 42

3.5 EMSA Analysis coupled to Western Blot Analysis 43

3.6 Supercoiling Assays . . . 43
3.7 *In Vitro* Transcription Assays with TBP, TFB and the TBP-TFB Fusion . . . 44
3.8 *In Vitro* Transcription Assays with Rpo13 . . . 45
3.9 Attempts at removing Rpo13 from the RNAP . . . 45
3.10 Cell Extract Preparation . . . 46
3.11 GST-Pulldown Assay . . . 46
3.12 Yeast-two Hybrid Screen . . . 46
3.13 Co-Immunoprecipitation Assays . . . 47
3.14 'Baby Machine' Cell Synchronisation Procedure . . . 48
3.15 Yeast Protein Cell Extract Preparation . . . 49
3.16 ChIP-Sequencing . . . 49
3.17 RNA Sample Preparation for RNA-sequencing . . . 50
3.18 Real-Time Q-PCR . . . 50

4 The Archaeal-Specific Rpo13 Subunit 52

4.1 Introduction . . . 52
4.2 Results . . . 55
4.2.1 Rpo13's Cloning and Purification . . . 55
4.2.2 Antibody Raising against Rpo13 . . . 57
4.2.3 Rpo13's Binding to DNA . . . 58
4.2.4 The Purification and Characterisation of Rpo13's Mutants . . . 62
4.2.5 Analysis of Binding to DNA of Rpo13's Mutants . . . 63
4.2.6 Rpo13's Binding Mode to DNA . . . 65
4.2.7 Rpo13's Topological Impact on DNA . . . 67
4.2.8 Rpo13's *in vivo* Association with the Archaeal RNAP within the *S. acidocaldarius* Cell Extract . . . 70
4.3 Discussion . . . 72

5 The Function of Rpo13 in the Context of Transcription 83

5.1 Introduction . . . 83

5.2 Results . . . 85

5.2.1 *In vitro* Function of the TBP-TFB Fusion . . . 85

5.2.2 Role of Rpo13 in Transcription . . . 87

5.2.3 Attempts at Removing Rpo13 from the Archaeal RNAP . . . 91

5.3 Discussion . . . 94

6 Interaction between RPC34 and the Single Archaeal RNAP 101

6.1 Introduction . . . 101

6.2 Results . . . 104

6.2.1 *In vitro* RPC34's Interaction with the Archaeal RNAP . . . 104

6.2.2 RPC34's Cloning and Purification . . . 106

6.2.3 Antibody against RPC34 . . . 107

6.2.4 *In vivo* RPC34 Interaction with the Archaeal RNAP . . . 108

6.2.5 RPC34's Interaction with the Archaeal TFB-2 . . . 114

6.3 Discussion . . . 120

7 Analysis of Rpo13 Genomic Distribution 125

7.1 Introduction . . . 125

7.2 Results . . . 127

7.2.1 Rpo13's *In Vivo* Genomic Localisation versus the One of the RNAP . . . 127

7.2.2 Transcript Abundance versus RNAP Localisation . . . 130

7.2.3 Origin Firing versus Transcript Abundance and RNAP Localisation . . . 135

7.3 Discussion . . . 139

8 Conclusions 145

Bibliography 150

Chapter 1

Introduction

1.1 The Archaeal Domain of Life

Living organisms have been classically divided into basic dichotomies. Traditionally, what was not considered to be part of the plant kingdom was considered to be part of the animal kingdom. The discovery of bacteria, which were found to be very different from animals and plants, led to the definition of Eukaryotes and Prokaryotes, where eukaryotic cells are thought to be the product of symbiotic events between primordial unicellular organisms and to contain higher order cellular organelles such as the nucleus. It was not until the 1970s that, for the first time, ribosomal RNA (rRNA) was used as a molecular chronometer since it is not only universal to all organisms but can also be easily isolated. In addition, the rRNA sequence changes slowly with time, where the rates of base substitution vary according to the part of the molecule, allowing for phylogenetic analyses. This led to the division of living systems into Eukaryotes, Bacteria and a new domain of life, Archaea (Figure1.1) [176], [178]. Indeed, whilst archaeal rRNAs were found to be comparable in size to bacterial ones, they share very little sequence similarity and patterns of base modification to bacterial rRNAs [176].

Archaea, despite being unicellular prokaryotic organisms and lacking the nuclear envelope similarly to bacteria, have very unique features, which are domain-specific.

These include distinct aspects of electron transport phosphorylation, lipid structure and biosynthesis, flagellin units, split of the Rpo1 subunit of the RNA polymerase (RNAP), RNAP gene order, unique modified nucleotides in tRNA and unique sequences flanking rRNA genes [184]. More importantly, they are a mosaic of bacterial and eukaryotic molecular features. For example, archaeal DNA replication is carried out by enzymes that are homologous to eukaryotic ones [52], [106]. Also, the archaeal RNAP is similar to the eukaryotic RNAP-II in both terms of subunit composition and structure ([171]), and the basal archaeal transcription initiation factors are similar to eukaryotic ones [114]. Furthermore, more than 30 ribosomal proteins are shared between the Archaea and Eukaryotes, and many archaeal translation factors are related to eukaryotic ones [105], [114]. Finally, some of the systems involved in operational functions, such as vacuolar ATPases ([63]) and secretion pathways ([65], [64]) are related to eukaryotic ones. On the other hand, several features are similar to bacteria, such as a circular chromosomal architecture ([165]), coupling between transcription and translation ([58]), capless translation initiation ([27]) and the cell division system [150], [109].

The evolutionary relationship between the Archaea and Eukaryotes is divided into two views [66] (Figure1.2). The first one proposes that Eukaryotes and Archaea are two distinct sister lineages derived from a common ancestor, where the features shared between the Archaea and Eukaryotes are ancestral traits that were inherited from their common ancestor. This view is reflected in the classical monophyletic universal tree of life, with the three primary domains of life. The second one proposes that Eukaryotes arose from the association of a bacterium and an archaeon. This view claims that the universal tree of life is composed of only two primary domains, the Archaea and Bacteria, where Eukaryotes represent a secondary domain that came into existence after the first two. Three studies based on protein analysis support the first view, while four studies based on genome content analysis support the second view [66]. The second view is most likely to be the most valid one as Rpo8 is found in the *Crenarchaeota*

but not in the *Euryarchaeota*. Phylogenetically, the Archaea fall into four main distinct phyla: the *Euryarchaeota*, the *Thaumarchaeota*, the *Crenarchaeota*, and the *Korarchaeota*, with possibly a fifth phylum, the *Aigarchaeota* (Figure1.1) [177], [2], [25], [82], [55]. The *Euryarchaeota* comprise methanogens, extreme halophiles, thermophiles, psychrophiles and sulphate-reducing species, while the *Thaumarchaeota* comprise ammonia-oxidizer species, and finally the *Crenarchaeota* comprise sulphur-dependent species and other thermophiles. Of the remaining phyla, only one member is known for the *Korarchaeota* ([55]), while the only known nanoarchaeum ([82]) is thought to be an early branching member of the *Euryarchaeota* [24]. Many Archaea are extremophiles, and are found in geological niches originally thought to be uncolonisable, such as hot springs, salt lakes and sulphurous caldrons as well as other more common yet extreme habitats such as soils, oceans, marshlands and the human colon.

In this particular study, the hyperthermophile *Sulfolobus acidocaldarius* DSM639 strain is the organism of choice, which is the first hyperthermoacidophile to be characterised from acid hot-springs at Yellowstone National Park, USA ([26]), and the name was chosen to describe the irregular lobe-shaped character of the organism. It grows optimally at 75 to 80 °C and at a pH of 2 to 3, under aerobic conditions and on organic substrates, such as yeast extract, tryptone, casamino acids and a limited number of sugars. It has special features, such as the ability to exchange chromosome genes intercellularly ([1], [67]), and to grow synchronously in culture ([20], [19]), enabling cell cycle studies, in addition to a very stable genome organisation. Its complete sequence is also avaliable [38]. Finally, and most importantly, its thermophilicity allows for *in vitro* biochemistry experiments, where most recombinant proteins are stable to heat-treatment. This study revolves around the *Sulfolobus* RNAP which, in addition to being similar to the eukaryotic RNAP as opposed to the bacterial RNAP, has unique and distinct features.

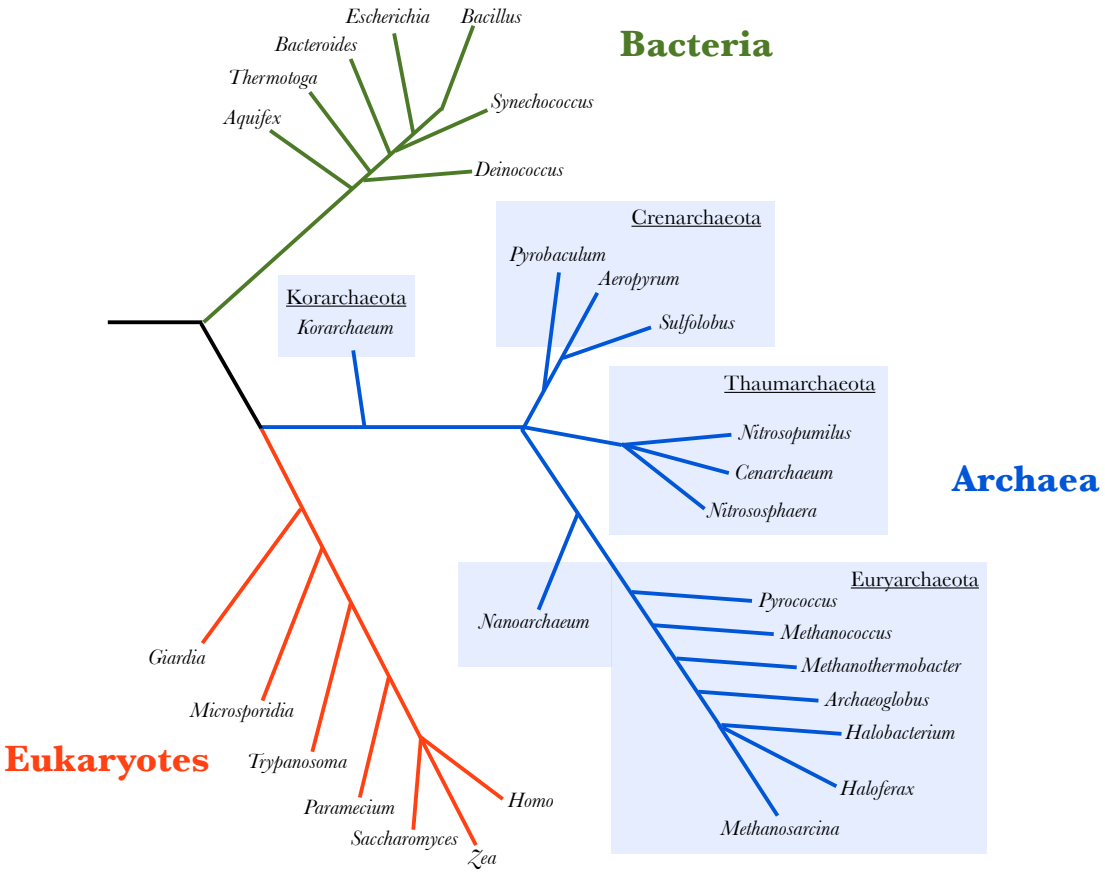


Figure 1.1: *The three of life is divided into three domains: the Eukaryotes, the Archaea and Bacteria. Each domain of life in turn comprises several phyla. Specifically, upon rRNA study, the domain Archaea has been divided into several phyla. The Euryarchaeota is the most diverse group, including methanogens, extreme halophiles, thermophiles, psychrophiles and sulphate-reducing species. The Thaumarchaeota comprise ammonia-oxidizer species, while the Crenarchaeota comprise sulphur-dependent species and other thermophiles. Finally, both the Nanoarchaeota and the Korarchaeota have each one known member only. The figure is recreated from [3].*

1.2 Transcription as a Cellular Process

Transcription is the process of creating a complementary RNA copy of a sequence of DNA. Being one of the core cellular processes, it is of not much surprise that RNAPs are highly conserved in evolution ([171]), where all existent RNAPs have a common structural framework and use the same basic molecular mechanisms. Only a single RNAP is present in Archaea and Bacteria, as opposed to the three RNAPs that direct eukaryotic RNA synthesis: RNAP I, RNAP-II and RNAP-III.

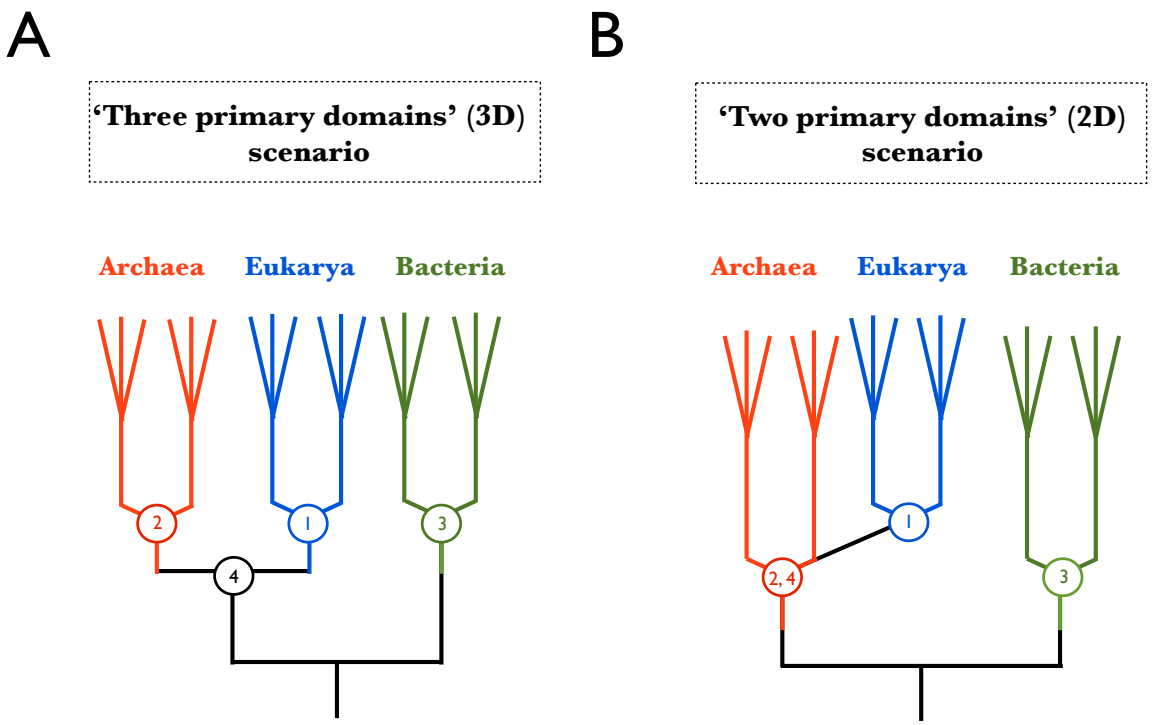


Figure 1.2: *Proposed models for the evolution of the three domains of life.* ***(A)*** *In the 'three primary domains' (3D) scenario, the Eukarya, Archaea and Bacteria form three distinct domains, each with a specific most recent common ancestor (1,2 and 3). The Eukarya and Archaea are two different sister lineages, where they share an ancestor (4), the nature of which remains undetermined. The branch lengths and number of lineages within each domain are arbitrary.* ***(B)*** *In the 'two primary domains' (2D) scenario, the Archaea and Bacteria are the two primary domains, while the Eukarya is a secondary domain that came into existence after the merging of an archaeon and a bacterium (for simplicity, only the relationship between the Archaea and the Eukarya is shown). In this case, the most recent ancestor of the Eukarya (1) derives directly from within the Archaea. Hence, the most recent ancestor of the Archaea is also the most recent ancestor of the Archaea and the Eukarya (2,4). The branch lengths and number of lineages within each domain are arbitrary. The figure is recreated from [66].*

While RNAP I and RNAP-III are in charge of the transcription of non-coding RNAs, such as all rRNAs with the exception of 5S rRNA for RNAP I, and 5S rRNA as well as tRNAs and other small RNAs for RNAP-III, RNAP-II directs the transcription of coding genes. Transcription is divided into initiation, elongation and termination. Steps during transcription initiation include promoter DNA binding, DNA melting, and initial synthesis of short RNA transcripts. The transition from initiation to elongation (promoter escape) leads to a stable elongation complex that is characterised by a transcription bubble. A multitude of transcription factors

are involved in directing transcription in its three stages of initiation, elongation and termination. It is important to point out that some transcription factors that are permanently incorporated in one RNAP system are on the other hand reversibly incorporated in another RNAP system. Therefore, the boundary between core RNAP subunits and associated transcription factors is loose in its definition. Much insight into the mechanisms of transcription is derived from crystallographic structures of the eukaryotic RNAP-II. Nonetheless, structures of the archaeal RNAP, which is very similar in structure and subunit composition to RNAP-II, have also been obtained. In the following section, the basic structural principles underlying transcription directed by RNAP-II will be reviewed, where the same basic principles are also valid for archaeal transcription despite slight differences.

1.3 Eukaryotic Transcription

The eukaryotic RNAPs are multisubunit enzymes, where RNAP I, RNAP-II and RNAP-III comprise 14, 12 and 17 subunits respectively. Ten subunits form a structurally conserved core, and further subunits are found in the periphery. Specifically, the RNAP-II core is made of Rpb1, Rpb2, Rpb3, Rbp5, Rpb6, Rpb8, Rpb9, Rpb10, Rpb11 and Rpb12 subunits and a peripheral heterodimer of Rpb4 and Rpb7 subunits (Figure1.3). The Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12 subunits are shared between RNAP I, RNAP-II and RNAP-III.

1.3.1 RNAP-II's Overall Structure

Large amounts of endogenous *Saccharomyces cerevisiae* RNAP-II protein preparations have been used for high resolution structural studies of the general RNAP-II backbone model, where initially a special *rpb4* deletion yeast strain was used in order to avoid heterogeneity due to the Rpb4/7 complex that could prevent crystallisation [48], [53]. The initial low resolution three-dimensional structure of yeast RNAP-II was determined by electron micrographs of two-

dimensional crystals ([47]), which revealed a channel of about 25 Å in diameter hypothesised to contain double-stranded (ds)DNA (Figure1.3) as well as a narrower channel that branches off from the 25 Å diameter channel. Higher resolution X-ray crystallographic structures were subsequently obtained from three-dimensional crystals, which revealed more detailed molecular features. The crystal structure obtained from Fu and coworkers showed that the channel was narrower (14 Å) and no longer sufficient to accommodate dsDNA, though large enough for the passage of single-stranded (ss)DNA or RNA [59]. A mobile element denoted as the clamp domain was also revealed (Figure1.3) [59]. The subsequent structures ([45], [46]) divided the RNAP-II into functional modules. The core module consists in Rpb1 and Rpb2, these subunits forming distinct masses with a deep positively charged cleft between them that contains the active site centre, previously identified as the channel (Figure1.3). Elements of Rpb1 and Rpb2 traverse the cleft, with the first active site magnesium ion (metal ion A) occurring within a prominent loop of Rpb1, which harbours the conserved aspartate residue motif loop involved in magnesium ion binding [45]. A second magnesium ion (metal ion B) is weakly bound between the Rpb1 aspartate loop and one or two conserved acidic residues in Rpb2. These two magnesium ions are thought to participate in a 'two-metal-ion' mechanism for the catalysis of phosphodiester RNA bond formation universal among all polymerases [154]. Beyond the active site, the DNA path is blocked by a protein wall (Figure1.3) [45]. While the clamp domain forms the Rpb1 side of the cleft, the Rpb2 side of the cleft consists of two domains, termed lobe and protrusion (Figure1.3). Thus, the active centre is formed by the floor of the cleft at its end, and is located between the protrusion, the wall and the clamp. Due to the presence of the protein wall, the DNA-RNA hybrid emanating from the active site would have to pass up the wall, at nearly right angles to the incoming DNA in the cleft. A hole in the floor of the cleft beneath the active site would allow entry of substrate nucleotide triphosphates [45]. The core module also comprises Rpb3, Rpb10, Rpb11 and Rpb12 subunits, which are involved in RNAP-II assembly and bridge one side of the the Rpb1 and Rpb

2 subunits. Around the periphery of the enzyme, Rpb5, Rpb6 and Rpb8 assemble with Rpb1, and Rpb9 binds to both Rpb1 and Rpb2. DNA is suggested to enter the cleft passing between a pair of mobile elements termed jaws (Figure1.3). The jaw lobe module is the upper jaw and comprises Rpb1, Rpb2 and Rpb9 subunits (Figure1.3) [46]. The shelf module is the lower jaw, and comprises Rpb1, Rpb5 and Rpb6 subunits (Figure1.3) [46]. Very importantly, residues in the Rbp5 loops facing the DNA are conserved. Two prolines present their side chains to the DNA with a spacing and relative orientation appropriate for contacting the DNA backbone [45]. Furthermore, the clamp domain, which is thought to permit entry of promoter DNA, comprises regions of Rbp1, Rpb6 and Rpb2 [45]. Finally, zinc ions near the RNAP-II surface appear to have a structural role and stabilise the enzyme. Indeed, while Rpb1 and Rpb9 each bind two zinc ions, subunits Rpb2, Rpb3, Rpb10 and Rbp12 each bind one zinc ion [46]. In addition, three zinc ions stabilise the unique clamp fold [46].

1.3.2 RNAP-II's Catalytic Elements

Structures of RNAP-II elongation complexes revealed the DNA path in the enzyme as well as structural elements involved in strand separation and transcription bubble maintenance [62], [175], [90]. Work by Westover and coworkers and Kettenber and coworkers yielded the elongating RNAP-II in the post-translocation state with the use of synthetic oligonucleotides ([175], [90]), with the nucleotide that was just added to the RNA having advanced to the -1 site, as opposed to the RNAP-II in the pre-translocation state obtained by Gnatt and coworkers obtained with the assembly of transcribing complexes [62]. The structure of transcribing RNAP-II was found to differ from the one of free RNAP-II mainly in the position of the clamp [62]. The clamp was found to swing over the cleft during the formation of the transcribing complex, trapping the template and transcript. The DNA was found to contact the Rbp5 jaw domain at the loop containing the proline residues, to then pass between the Rpb1 clamp head to finally reach the bridge helix and bind to switches

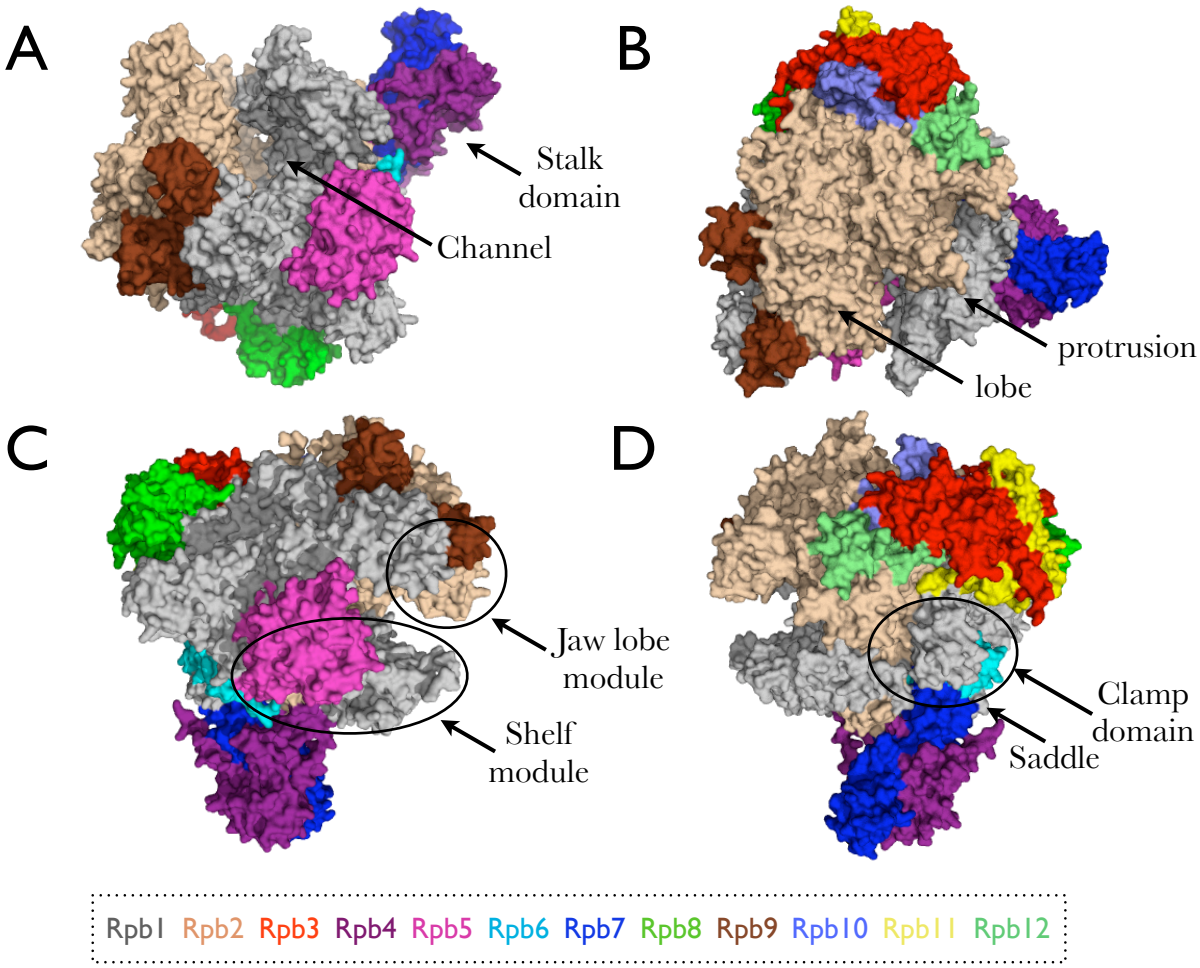


Figure 1.3: *The eukaryotic RNAP-II is composed of 12 subunits, Rpb1-12, where each subunit is colour-coded. Four different stereo-views of the RNAP-II are shown (Figure A-D), where functional modules are indicated by arrows.* ***(A)*** *The RNAP-II channel traverses the polymerase, and consists of the cleft containing the active site. A protein wall is present, which modificates the DNA path is a way to bend the DNA at nearly 90°. The active site is therefore surrounded by the wall, clamp and protrusion domains. The stalk domain is composed of Rpb4 and Rpb7 subunits.* ***(B)*** *The Rpb2 subunit is defined in terms of lobe and protrusion domains.* ***(C)*** *The DNA enters the channel via gripping through jaw modules. The upper jaw module is called the jaw lobe module and is composed of Rpb1, Rpb2 and Rpb9 subunits. The lower jaw module is called the shelf module and is composed of Rpb1, Rpb5 and Rpb6 subunits.* ***(D)*** *The clamp domain closes onto the DNA, and is composed of Rpb1, Rpb6 and Rpb2 subunits. The saddle is part of the RNA exit groove. The PDB code for the RNAP-II crystal structure is 3FKI.*

1 and 2 (Figure1.4) [62]. Downstream DNA unwinding is achieved through three positively charged residues in switch 2, which pull the template strand away from the duplex axis, and by three negatively charged residues in switch 1 that repel the DNA strand [90]. Also, the Rpb2 fork loop 2 sterically blocks duplex binding,

interferes with the non-template strand and prevents re-association of separated strands, while the Rpb2 fork loop 1 interacts with RNA and may serve to delimit the region of RNA-DNA strand separation, maintaining the transcription bubble further upstream (Figure1.4) [62], [175], [90]. Furthermore, three loops extending from the clamp, termed rudder, lid and zipper have been suggested to play roles in hybrid dissociation, RNA exit and maintenance of the upstream end of the transcription bubble [62]. The lid is involved in maintaining the separation of RNA and template DNA strands ([175]), while the zipper may play a similar role in separating template and non-template DNA strands (Figure1.4). The rudder is not directly involved in strand separation, but rather stabilises unwound DNA [175]. Finally, extension of RNA leads to the exit groove, which is created by the saddle, lip and flap loop, and nucleotide triphosphates enter through the pore during RNA synthesis (Figure1.4) [62]. The DNA path at the active centre displayed a right angle bend, which orients the DNA-RNA hybrid optimally for transcription. Hence, the DNA would not pass through the channel as initially proposed, but rather would be associated with the RNAP-II entirely on one face [135].

1.3.3 RNAP-II's Rpb4/Rpb7 Complex

Upon crystal structure resolution of the archaeal Rb4/Rpb7 stalk complex on its own (Figure1.3) ([158]), Craighead and coworkers were able to derive information from the Rb4/Rpb7 complex in order to obtain the crystal structure for the entire RNAP-II [44]. The Rb4/Rpb7 complex was proposed to interact with RNAP-II through Rpb4 [44]. Also, the position of the Rbp4/Rpb7 complex, adjacent to the RNA exit groove, suggests that two putative ss-nucleic acid binding domains in Rpb7, the ribonucleoprotein (RNP) fold and oligonucleotide-binding (OB) fold, are likely involved in determining the path of the newly synthesised RNA as it is transcribed, as shown both *in vitro* and *in vivo* [44], [121], [159], [36], [6]. Work by Asturias and coworkers hinted for a RNAP-II structure with a lack of protein density on the cleft ([9]), that was previously observed to clamp DNA in the cleft

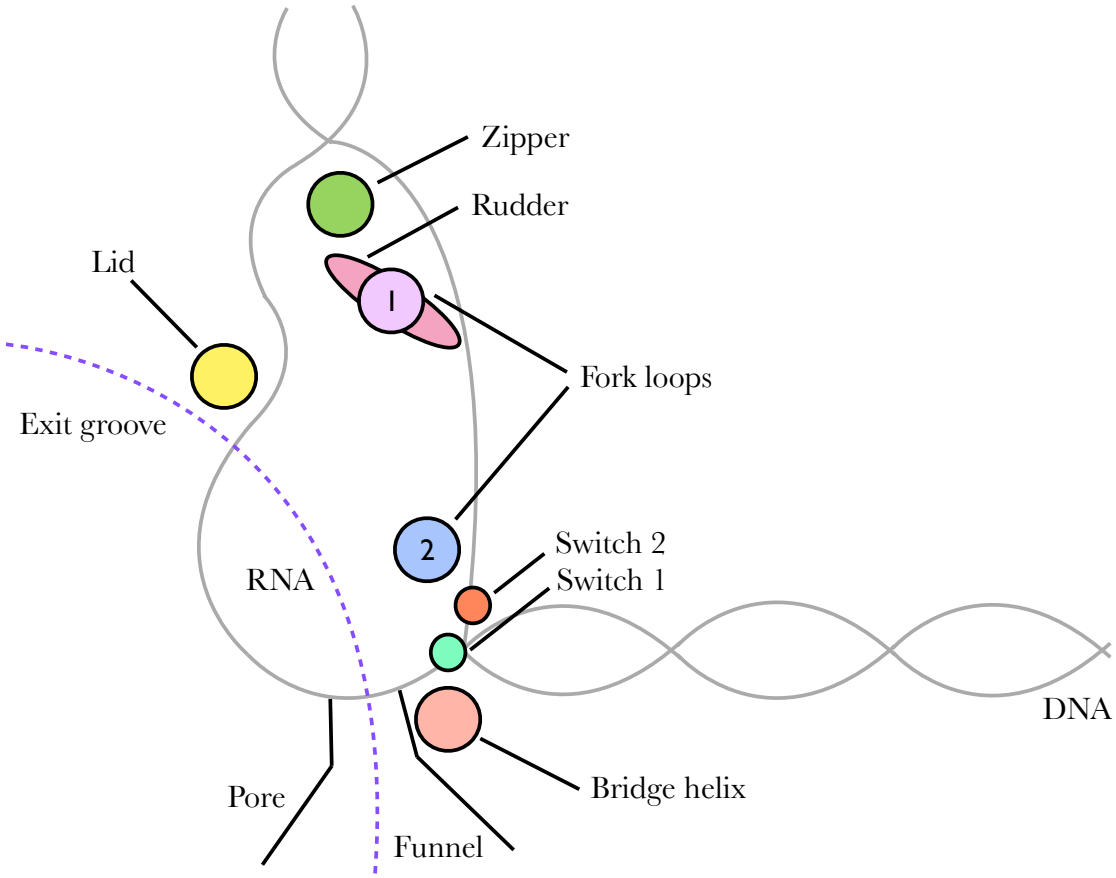


Figure 1.4: *Several key catalytic elements are involved in strand separation and transcription bubble maintenance. Upon DNA entry into the RNAP-II through the DNA-gripping jaw domains and clamp domain, the DNA helix is unwound by switches 1 and 2 and it encounters the bridge helix, which is involved in the catalysis of nucleotides and RNAP-II translocation. Furthermore, fork loop 2 mainly prevents re-association of unwound DNA strands, while fork loop 1 mainly maintains the transcription bubble further upstream. Moreover, while the zipper and lid are respectively involved in the separation of DNA strands and DNA-RNA strands, the rudder is not directly involved in strand separation but rather it stabilises unwound DNA. Finally, while the RNA exits through the exit groove, nucleotide triphosphates enter through the funnel-shaped pore for RNA synthesis. The figure is recreated from [62].*

[47]. They suggested this 'open conformation' of RNAP-II, where association of Rpb4 and Rpb7 subunits RNAP-II favours on the other hand the other 'closed conformation'. Indeed, subsequent work by Armache and coworkers revealed that not only the Rpb4/Rpb7 complex binds to the RNAP-II through Rpb7 (Figure1.3), but also that modeling the clamp in the 'open conformation' results in a clash with the Rpb7 tip [8]. Also, the structure of the complete RNAP-II in the clamp-closed

state showed that loops in Rpb1 that form the hinge for clamp movement are at the level of the RNA exit groove, where contacts of Rpb4 above the groove and Rpb7 below the groove would appear to bracket the clamp, constraining it in the closed state [33]. Therefore, the Rpb4/7 complex can only bind when the clamp is closed, and the clamp can only open after the dissociation of the Rpb4/Rpb7 complex from the RNAP-II core. The requirement for Rpb4/Rpb7 for the inititation of transcription ([54]), with its effect on the clamp constraining it in the closed state, suggests that promoter DNA binding and transcription initiation occur in the clamp-closed state with bending of the DNA. Specifically, the DNA can either be melted in the cleft with expelling of the DNA non-template strand before clamp closure, or the DNA can be melted on top of the cleft, with slipping of the template strand into the cleft. As further discussed in section 'Transcription Initiation by RNAP-II', the second model is favoured. It has to be noted that the eukaryotic Rpb4/Rpb7 is permanently associated with the RNAP and plays an important role during all stages of the transcription cycle, as shown by ChIP-seq data [86].

1.3.4 Catalysis and Transcription Elongation by RNAP-II

Mechanisms for nucleotide selection and catalysis as well as transcription elongation have been proposed. In order to investigate the mode of interaction with the nucleotide, the RNAP-II structure in the post-translocation state was obtained upon soaking of crystals with both the nucleotide matched to the upstream DNA base and unmatched nucleotides [174]. Two nucleotide binding sites were found (A and E sites), where the nucleotide would first enter the E site and then rotate around the magnesium ion (metal ion B) to sample base pairing in the A site, where the template strand is held fixed by interaction with the bridge helix [174]. Furthermore, contacts between the bridge helix, trigger loop and the nucleotide in the A and E sites occur [168]. The trigger loop, which is located beneath the nucleotide in the A site, refolds into two trigger helices upon nucleotide binding that, together with the bridge helix, creates a three-helical bundle [163], [42]. The

trigger loop on its own as well as its interactions with the bridge helix and with other Rpb1 and Rpb2 residues lining the active centre region serve to recognise all features of the nucleotide in the A site and parameters of the DNA-RNA hybrid helix in such a way to discriminate against base-pair mismatching [168], [42], [88], [42]. Thus, the interactions between the trigger loop and the bridge helix bend the latter in the direction of the nucleotide and the DNA-RNA hybrid, consistent in the involvement of the trigger loop in the control of bridge helix motion [168]. Upon correct nucleotide incorporation and pyrophosphate release, the interaction between the incorporated nucleotide and the refolded trigger helix is destabilised and the DNA-RNA hybrid is freed for translocation, where the new RNA 3' end is aligned with the magnesium ion (metal ion A) [168]. Thus, movement of the trigger loop coupled to that of the bridge helix may contribute to the translocation process. Indeed, the use of the elongation inhibitor α-amanitin revealed that α-amanitin traps the trigger loop in a new conformation with the bridge helix, stabilising the RNAP-II in a translocation intermediate [28]. Furthermore, the same protein-nucleic acid interactions formed at the upstream and downstream edges of the DNA-RNA hybrid and the downstream DNA are present in the pre-translocated and post-translocated states, suggesting that, in the absence of nucleotide, these two states are equally stable [162]. The structure therefore supports a 'brownian-ratchet' mechanism, in which translocation is driven by thermal motions giving rise to equilibrium between the post-translocated and pre-translocated states of the nucleotide-free elongation complexes, while binding of the nucleotide would favour the post-translocated state via the trigger loop (Figure1.5).

1.3.5 Transcription Initiation by RNAP-II

Recent effort in the field has been redirected in understanding the structural basis of transcription initiation. Several transcription factors are involved in the formation of the pre-initiation complex (PIC) on promoter DNA: TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH. TFIID's association with the TATA box in promoter DNA or

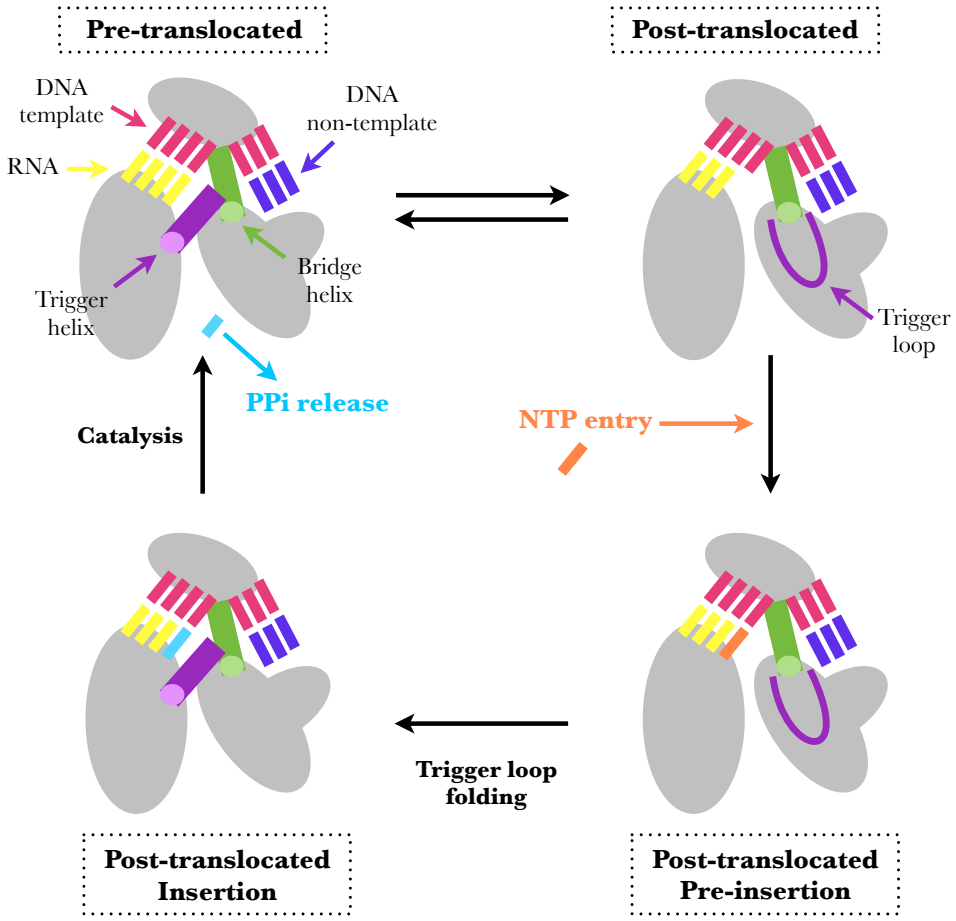


Figure 1.5: *A 'brownian-ratchet' mechanism is proposed for the RNAP-II translocation upon phosphodiester RNA bond formation. In the absence of nucleotide, the pre-translocated and post-translocated states are equally stable. Upon correct nucleotide binding in the A site, the trigger loop, which is located beneath the nucleotide in the A site, refolds into two trigger helices which, together with the bridge helix, creates a three-helical bundle. The post-translocation state is also favoured versus the pre-translocation state. Interactions between the trigger loop and the bridge helix occur in a manner to bend the bridge helix in the direction of the nucleotide and the DNA-RNA hybrid. These interactions also serve to detect all features of the nucleotide in the A site and parameters of the DNA-RNA hybrid helix in such a way to discriminate against base-pair mismatching. Upon correct nucleotide incorporation and pyrophosphate release, the interaction between the incorporated nucleotide and the refolded trigger helix is destabilised and the DNA-RNA hybrid is freed for translocation into the pre-translocated state for another cycle of nucleotide incorporation. The figure is recreated from [163].*

other elements such as the initiator sequence in TATA-less promoters requires the preceding TFIIB's recruitment, since TFIID is incorporated through contacts with both TFIIB and the TATA box, where TFIIA can stabilise the TFIID-DNA complex, and creates a context for TFIIB binding. TFIIB also binds to the TFIIB response

element (BRE) promoter DNA element. Analysis by surface plasmon resonance showed that TFIIB, TFIIE and TFIIF bind to the RNAP-II, while TFIIH binds to TFIIE [32]. TFIIF is involved in the prevention of non-specific DNA binding, while TFIIE and TFIIH are involved in DNA melting. A stepwise assembly model has been proposed for transcription factor incorporation where TFIID, TFIIB and TFIIF are minimally required for loading of promoter DNA onto RNAP-II, with the RNAP-II-TFIIF complex binding to the TBP-TFIIB-DNA complex [30], [103], [41]. As an alternative to the stepwise assembly of the initiation complex, it has been suggested that a large RNAP-II holoenzyme can be recruited to the promoter in a single step [95], [131]. Such holoenzyme was purified from yeast and mammalian cells, and comprise the RNAP-II as well as transcription factors [95], [131]. The various crystal structures containing transcription factors in the presence of DNA and/or RNAP-II as well as complementary experiments such as cross-linking protein-protein and protein-DNA analysis are described in the following paragraphs.

The TATA box binding protein (TBP), along with several TBP-associated factors that make up TFIID, is involved in DNA bending. The C-terminal core of TBP is highly conserved and contains two repeats that produce a saddle-shaped structure that straddles the DNA (Figure1.6) [129], [35]. This region binds to the TATA box, where contacts with the DNA minor groove are made with a β-sheet (Figure1.6), and interacts with transcription factors and regulatory proteins [129], [35]. By contrast, the N-terminal region varies in both length and sequence among different organisms [129], [35]. More specifically, in the TBP-DNA co-crystal, TBP-DNA contacts are completely restricted to the phosphate-ribose backbone and the minor groove [92]. Interactions with the minor groove can be divided into three classes [92]. Firstly, at either end of the TATA-box there are two pairs of phenylalanine side chains partially inserted between adjacent base pairs, producing the two dramatic kinks. Secondly, four polar side chains make minor groove H-bonds with acceptors of four base pairs centred about the approximate two-fold symmetry axis of TBP. Thirdly, fifteen residues projecting from the concave surface of TBP

make hydrophobic or van der Waals side-chain/base contacts with the eight-base-pair sequence flanking the approximate two-fold axis. A co-crystal structure with a longer TATA-box containing oligonucleotide showed that the entire TATA-box is severely bent toward the major groove, exposing the minor groove that interacts with the under-surface of the TBP-saddle [94]. Further studies have shown that there is a correlation between the degree of bending and the stability of the complex in solution, such that complexes with a longer lifetime appear to have increased bend angles [152]. Therefore, DNA bending is important for site-specific recognition by TBP. Furthermore, the TATA-box is centred on the pseudodyad of TBP, and is oriented so that the beginning of the TATA-box contacts the second direct repeat [94]. In response to the severe bend, the TATA-box unwinds considerably enabling the DNA to maintain nearly symmetrical contact with the under-surface of the TBP's saddle [94]. Thus, an induced-fit mechanism occurs upon DNA binding by TBP, where conformational changes in both the protein and the DNA generate complementary interaction surfaces without bad steric contacts between the DNA and protein atoms. TBP's overall role therefore consists in DNA binding and unwinding, assisting RNAP-II's start of initiation.

The TBP-TFIIB-DNA ternary complex crystal structure showed that ternary complex formation causes a small conformational change in TBP, produced by a twisting motion of one domain relative to the other [128]. In this complex, the rotational displacement of the two domains is intermediate between DNA-bound and apo-TBP. TFIIB was found to act as a clamp, binding to TBP in its cleft and interacting with the backbones of the coding and non-coding DNA strands upstream and downstream of the TATA box [128]. There are no contacts between TFIIB and major or minor groove edges of bases within the TATA-box element; rather, the basic surface of TFIIB interacts with the negatively charged phosphodiester DNA backbones [128]. Thus, TFIIB enhances the affinity of TBP for the core promoter. In addition, TFIIB binds to the BRE promoter DNA element, making major groove contacts with the fourth and fifth α-helices, which intercalate with

the DNA [110]. More importantly, TFIIB's central role is to recruit RNAP-II to the promoter DNA. The crystal structure of TFIIB complexed with RNAP-II not only elucidated TFIIB's structure, but also revealed the mode of interaction between TFIIB and RNAP-II [98], [113]. TFIIB was found to have an N-terminal zinc ribbon domain (B-ribbon), and a C-terminal domain (B-core, comprising two cyclin folds), while the connecting region was divided into the B-linker region, which comprises a helix and a β-strand, and a B-reader region, which comprises a loop and a helix (Figure1.6) [98], [113]. The B-ribbon binds to RNAP-II, while the B-core binds to DNA and TBP [98]. In the TFIIB-RNAP-II crystal structure, TFIIB extends from the dock domain (B-ribbon) via the RNA exit tunnel to the hybrid-binding site and active centre (B-reader), to the rudder and clamp (B-linker), and to the cleft's protein wall (B-core) (Figure1.7) [98]. TFIIB does not interact with Rpb4/7 in the structure, suggesting that the requirement of Rpb4/7 for initiation stems from a role in stabilising the clamp, which binds to the TFIIB's linker [98]. TBP was modeled on the other hand on the RNAP-II upstream face [98].

In order to obtain the closed promoter complex, the TFIIB-TBP-DNA complex was aligned with the TFIIB-RNAP-II crystal structure [98], [113]. It was found that downstream DNA runs between the clamp and the protrusion along the edges of the cleft towards the jaws, while upstream DNA extends along the Rpb2 side, where promoter DNA is initially recruited to RNAP-II by the B-ribbon that binds to the dock domain (Figure1.7)[98]. The DNA path is in agreement with the results obtained by hydroxyl radical cleavage of bound PIC to promoter DNA [123]. In order to obtain the open promoter complex, the DNA from the elongation complex structure ([126]) was placed in the TFIIB-RNAP-II crystal structure ([98]). It was found that the point of DNA opening lies above a tunnel lined by the B-core, B-linker, B-reader and the RNAP-II protrusion, fork loop 1 and rudder (Figure1.7) [98], [113]. Thus, the transition from the closed to the open promoter complex involves a rotation of the downstream DNA duplex and slippage of the template strand into the template tunnel and active centre, probably aided by movement of the flexible

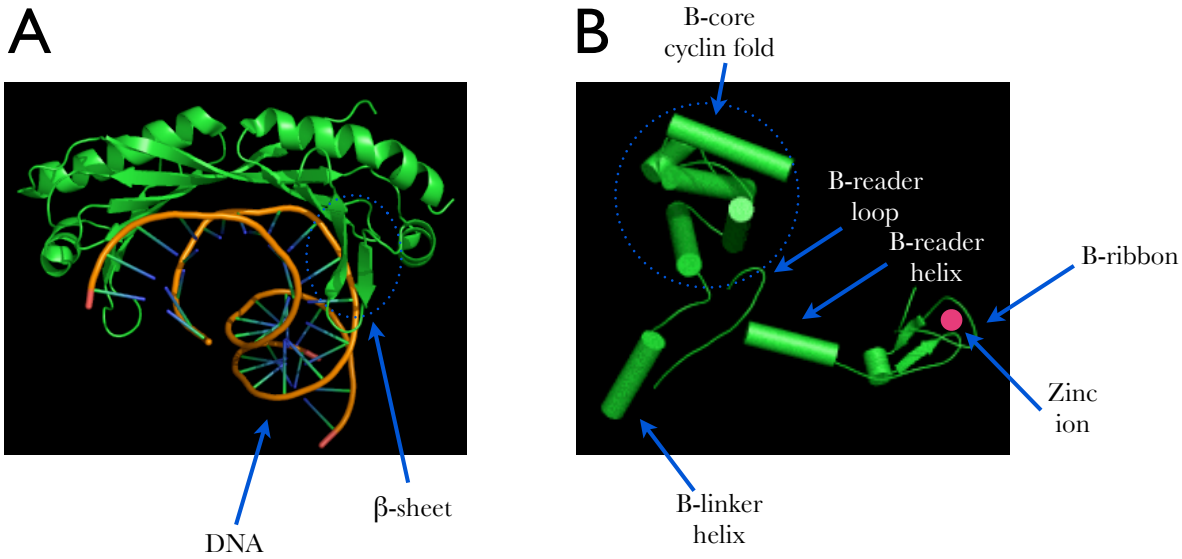


Figure 1.6: *Two main eukaryotic transcription factors are involved in the initial recruitment of RNAP-II onto DNA, TBP and TFIIB.* ***(A)*** *TBP displays a two-domain DNA-binding fold, and the protein's approximate two-fold intramolecular symmetry generates a saddle-shaped structure dominated by a curved antiparallel β-sheet, forming a concave DNA-binding surface. Contacts with the DNA minor groove are made with a β-sheet.* ***(B)*** *TFIIB is subdivided into four proteins regions. The C-terminal domain consists in the B-core, which comprises two cyclin folds, the N-terminal cyclin fold being shown in the figure, while the C-terminal cyclin fold is not resolved. The B-linker comprises a helix and a β-strand, while the B-reader comprises a loop and a helix. Finally, the N-terminal domain consists in the B-ribbon, which is a zinc ribbon domain. The PDB code for the TBP-DNA crystal structure complex is 1RM1, while the PDB code for the TFB crystal structure is 3K1F.*

fork loop 1 (Figure1.7). This model was subsequently confirmed, where crystal structures revealed that downstream DNA resides in the cleft as in the elongation complex, while the template strand extends over the bridge helix into the active centre generally along the path it takes in the elongation complex [42], [112]. Also, the B-linker is located between duplex DNA in the closed promoter complex and the template single strand in the open promoter complex, suggesting that it is involved in promoter opening [98]. DNA melting starts at around 20 base pairs downstream of the TATA box [98]. After open complex formation, the RNAP-II scans the DNA for an initiator (Inr) sequence motif that defines the transcription start site (TSS). Scanning involves threading of the template strand through the template tunnel along the active site, which is flanked by the B-reader, until an Inr sequence is

detected by the B-reader (Figure1.7) [98]. Indeed, in the open promoter complex, a conserved Inr T residue is adjacent to the B-reader helix, while the B-reader loop recognises the beginning of the Inr sequence and stabilises the first two nucleotides in the active centre [98]. The B-reader loop also contains the invariant Arg78 residue that may reach near the active site, which together with other conserved residues such as Glu62, once mutated, lead to TSS shifts [98]. RNA growth results in short DNA-RNA hybrids that are transiently bound by RNAP-II and TFIIB, leading to frequent ejection of short RNAs due to interference of the growing RNA with the B-reader loop [98], [113]. Promoter escape occurs when the transcript is at least seven nucleotides and long enough to form a stable RNA-DNA duplex of eight to nine base pairs, upon which TFIIB is released [98] [78].

The TBP-TFIIA-DNA ternary complex crystal structure showed that the primary interaction of TFIIA and TBP is between two parallel β-strands, which results in a 16-strand β-sheet in the complex [156]. TFIIA binds DNA exclusively through phosphate backbone contacts made within the TATA box and with bases upstream of it [156]. Also, the orientation of TBP with respect to that of the DNA was found to have changed by slipping two base pairs downstream along the DNA, as compared to the TBP-DNA binary complex [156].

In the RNAP-II-TFIIB-TFIIE crystal structure, the location of TFIIE suggests a direct role in events involving the arm around the cleft, where it may promote the closure of the arm by bridging the remaining gap [108]. Indeed, cross-linking results and EM data showed that the Rpb1's clamp domain interacts with TFIIE (Figure1.8) [37], [74]. TFIIE makes also extensive interactions with promoter DNA just downstream of the TATA box, and with and downstream of the transcription-bubble region, thus facilitating ss-DNA formation [123], [80].

Comparison of the RNAP-II-TFIIF crystal structure with the one of RNAP-II alone suggests a rearrangement of the RNAP-II induced by binding of TFIIF [43]. Specifically, binding of TFIIF to RNAP-II causes a rotation of the shelf and jaw lobe module and pushes apart the Rpb1 and Rpb8 feet ([43]), possibly through

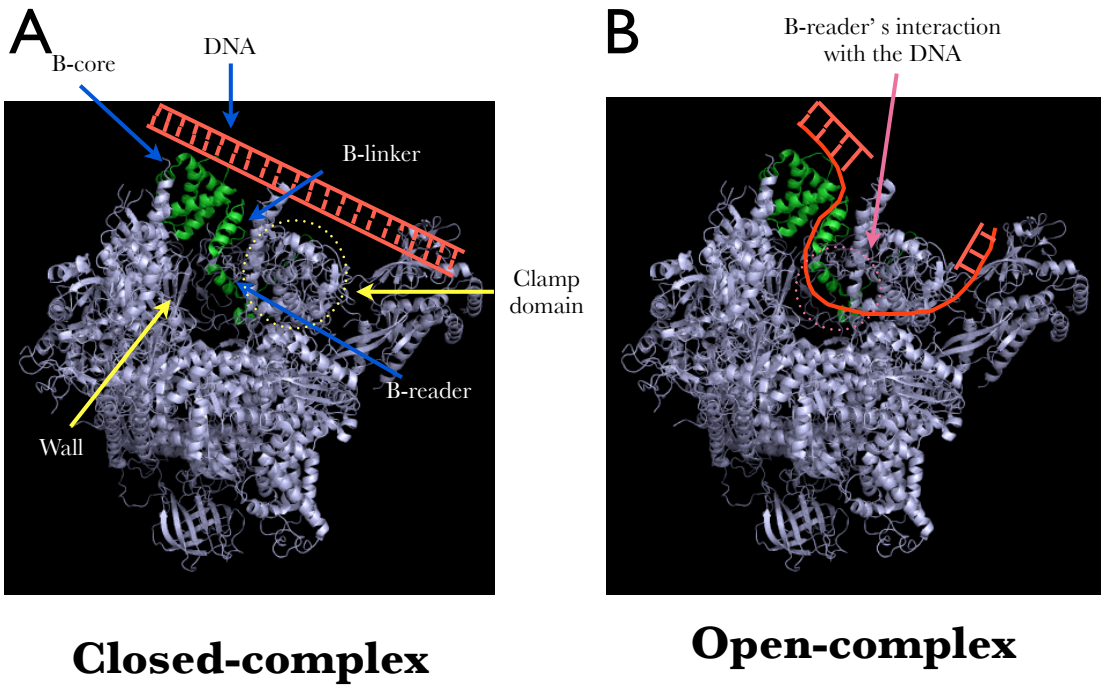


Figure 1.7: *TFIIB (green) extends into the active centre of the RNAP-II (grey) with its B-reader domain from the B-ribbon, to the rudder and clamp domain with its B-linker, and to the cleft's protein wall with its B-core.* ***(A)*** *Modeling of the closed-complex revealed that the DNA runs over the RNAP-II, where the DNA runs over the clamp and the protrusion along the edges of the cleft.* ***(B)*** *The point of DNA opening in the open-complex lies above a tunnel lined by the B-core, B-linker, B-reader and the RNAP-II protrusion. The transition from the closed to the open complex involves a rotation of the downstream DNA duplex and slippage of the template strand into the template tunnel and active centre. An Inr sequence is scanned for by the B-reader as the template strand is threaded into the tunnel. The figure is recreated from [98] and [112], where the PDB code for the RNAP-TFIIB crystal structure complex is 3K1F.*

association with Rpb1, Rpb2 and Rpb9 (Figure1.8) ([183], [74]), which is part of the jaw lobe module. A portion of TFIIF is also closely associated with the active site cleft and with the clamp, explaining the function of TFIIF in modulating interactions between RNAP-II and DNA [43], [37]. Furthermore, TFIIF interacts with TFIIB to affect TSS selection [57]. Finally, TFIIF was found to interact with promoter DNA both upstream and downstream of the TATA box ([123]), and to promote efficient promoter escape [181].

TFIIH is the responsible factor for Rpb1 phosphorylation, promoter melting, and promoter clearance. The role of TFIIH in promoter melting is to melt about one turn

of DNA encompassing the transcription start site to yield the transcription bubble. This process requires ATP and is mediated by the XPB subunit of TFIIH, which, in isolation, exhibits 3'-5' DNA-helicase activity. XPB acts as a molecular wrench interacting with downstream DNA to generate the torque that nucleates formation of the transcription bubble, where the presence of Rpb1, Rpb2 and Rpb5 on adjacent faces of the downstream DNA segment would preclude XPB from rotating by one full turn in a single step [93]. Therefore, XPB acts as a ratchet wrench, affecting rotation by one full turn in multiple smaller increments [93]. TFIIF, TFIIE or both would next insert residues between ds-promoter DNA to initiate and stabilise ss-bubble formation. This would be followed by insertion of the ss-template strand into the RNAP-II cleft. Perhaps promoted by the helicase activity of XPB, ss-template DNA would be fed through the RNAP-II cleft until a suitable TSS was encountered. TFIIH also stimulates efficient promoter escape, which requires ATP too and is again mediated by XPB [78]. Furthermore, in addition to promoter DNA binding, TFIIH interacts with the stalk domain, at the site of interaction with TFIIE (Figure1.8) [93], [74]).

Finally, the carboxy-terminal domain (CTD) of RNAP-II consists of up to 52 repeats of the sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser [120]. Other proteins often bind the CTD of RNAP-II in order to activate polymerase activity. It is the protein domain that is involved in the initiation of transcription, the capping of the RNA transcript, and attachment to the spliceosome for RNA splicing [151].

1.3.6 Other Eukaryotic Polymerases

RNAP I and RNAP-III share a structurally conserved core with RNAP-II, with the Rpb4/7 related heterodimeric subcomplexes A14/43 ([134], [100], [21]) and C17/25 respectively [60], [85], [56], [115]. In addition, RNAP I contains the subcomplex A49/34.5, while RNAP-III contains subcomplexes C37/53 and C82/34/31[56]. In particular, RPC34 has been proposed to be related to the TFIIE-β subunit (Figure1.9), in that it contains two adjacent winged helix (WH) domains in the

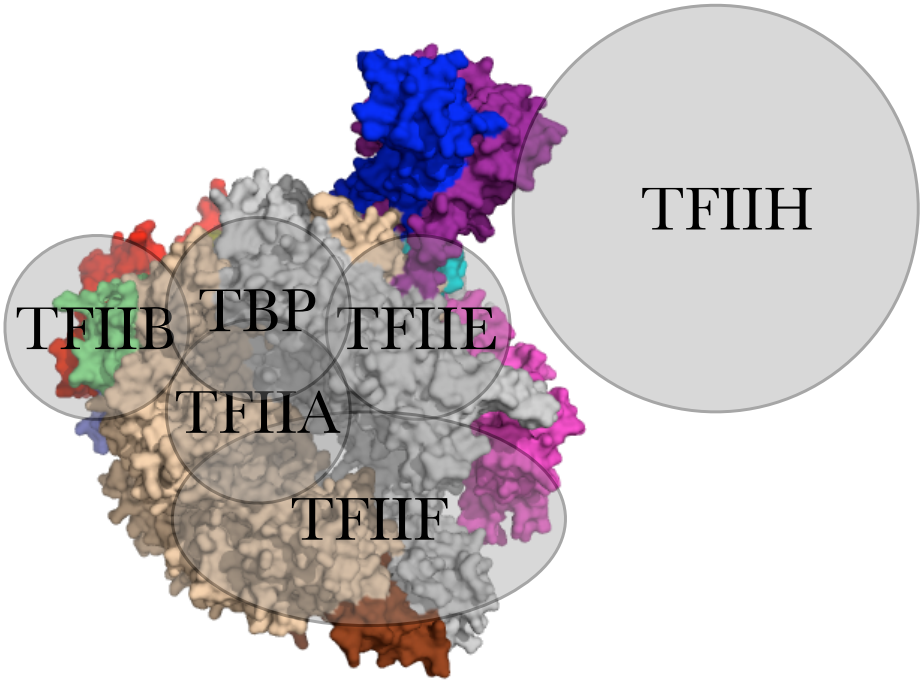


Figure 1.8: *EM data has revealed the binding sites for TBP, TFIIB, TFIIA, TFIIE, TFIIF and TFIIH. The in vitro reconstitution of this process suggests a sequential assembly pathway model for transcription initiation. TFIID, which comprises TBP, is the first factor which is recruited to the promoter. TFIIA and TFIIB are then recruited, further stabilising the interaction between TBP and promoter DNA. Subsequently RNAP-II, most likely in association with TFIIF, adds to the growing RNAP-II pre-initation complex. Finally, TFIIE and TFIIH, which is required for DNA melting, are recruited to form the transcriptionally competent pre-initation complex. The figure is recreated from [74].*

N-terminal region similarly to TFIIE-β, which contains WH domains [61]. Also, the WH domains in RPC34 are located on the clamp over the cleft similarly to TFIIE ([161]), where it has been proposed that RPC34 has a function in DNA opening (Figure1.9) [29]. Finally, TFIIB-like factors are also used, where TBP has a conserved role in binding TFIIB-like factors. For example, RNAP-III initiation requires the TFIIB-related factor 1 (Brf1 or TFIIIB70), which together with TBP and TFIIIB90 forms the transcription factor TFIIIB [89]. TBP also interacts with TFIIIC and SNAPc, which are required for recruitment of RNAP-III and TFIIIB at promoters. Importantly, while TBP and TFB directly recruit RNAP-II to the pre-initiation complex, RNAP-III requires the RPC34 subunit to mediate the interaction between TFIIIB70 and RNAP-III ([29], [91], [173]), where both the N-terminal

and C-terminal region of TFIIIB70 contribute to RPC34 binding [91], [5]. Also, the interaction between RPC34 and TFIIIB70 was found to be necessary for open promoter complex formation [29].

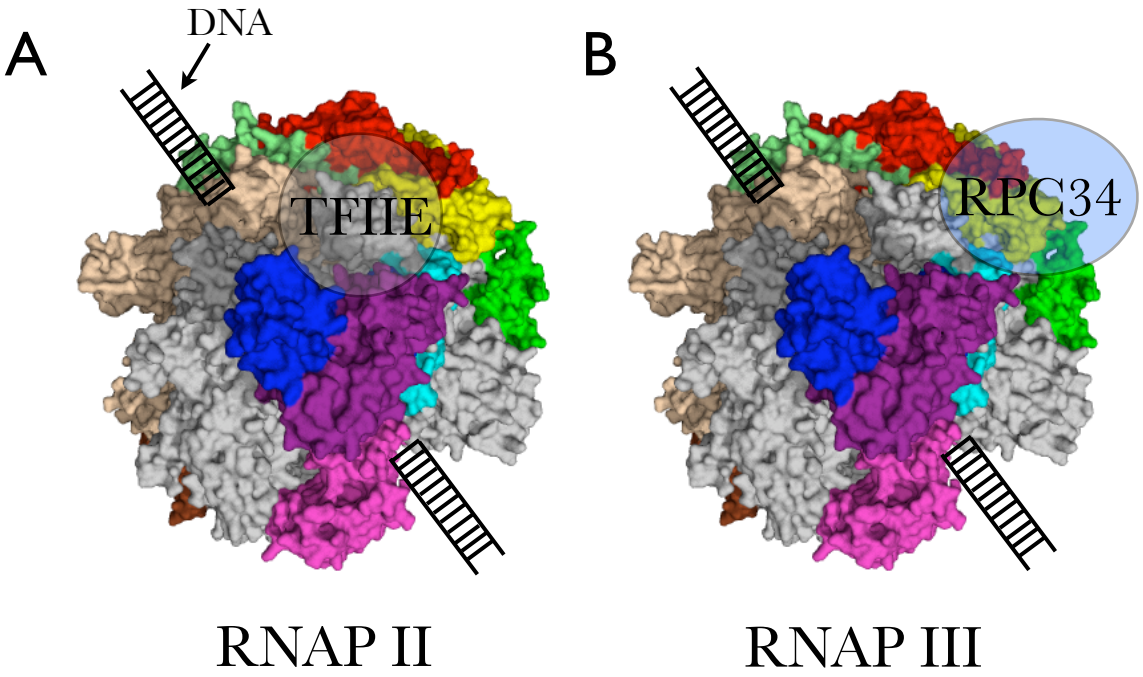


Figure 1.9: *Locations of TFIIE on RNAP-II and the TFIIE-related RNAP-III-specific RPC34 factor on RNAP-III.* ***(A)*** *TFIIE is located on the clamp domain on RNAP-II.* ***(B)*** *RPC34 is known to interact directly with subunit Brf1 of TFIIIB, where it is hypothesised that the RPC34 subunit is located to the RNAP-III saddle and stalk, allowing Brf1 to link the stalk subunits and the C82/C34/C31 subcomplex. The figure is recreated from [160], where the PDB code for the RNAP-II crystal structure is 3FKI. The RNAP-II protein backbone is used to also represent RNAP-III, as RNAP-III structures obtained in complex with C17/25, C37/53 and C82/34/31 are RNAPII-based homology models.*

1.4 Archaeal Transcription

As opposed to the eukaryotic three RNAPs, only a single RNAP is found in Archaea. It is similar in structure and subunit composition to RNAP-II, as opposed to the bacterial RNAP. The *Sulfolobus* RNAP contains 13 subunits, with a core of Rpo1N, Rpo1C, Rpo2, Rpo3, Rpo5, Rpo6, Rpo8, Rpo10, Rpo11, Rpo12 and Rpo13 subunits, and a peripheral heterodimer of Rpo4 and Rpo7 subunits (Figure1.10). It also lacks the CTD of RNAP-II.

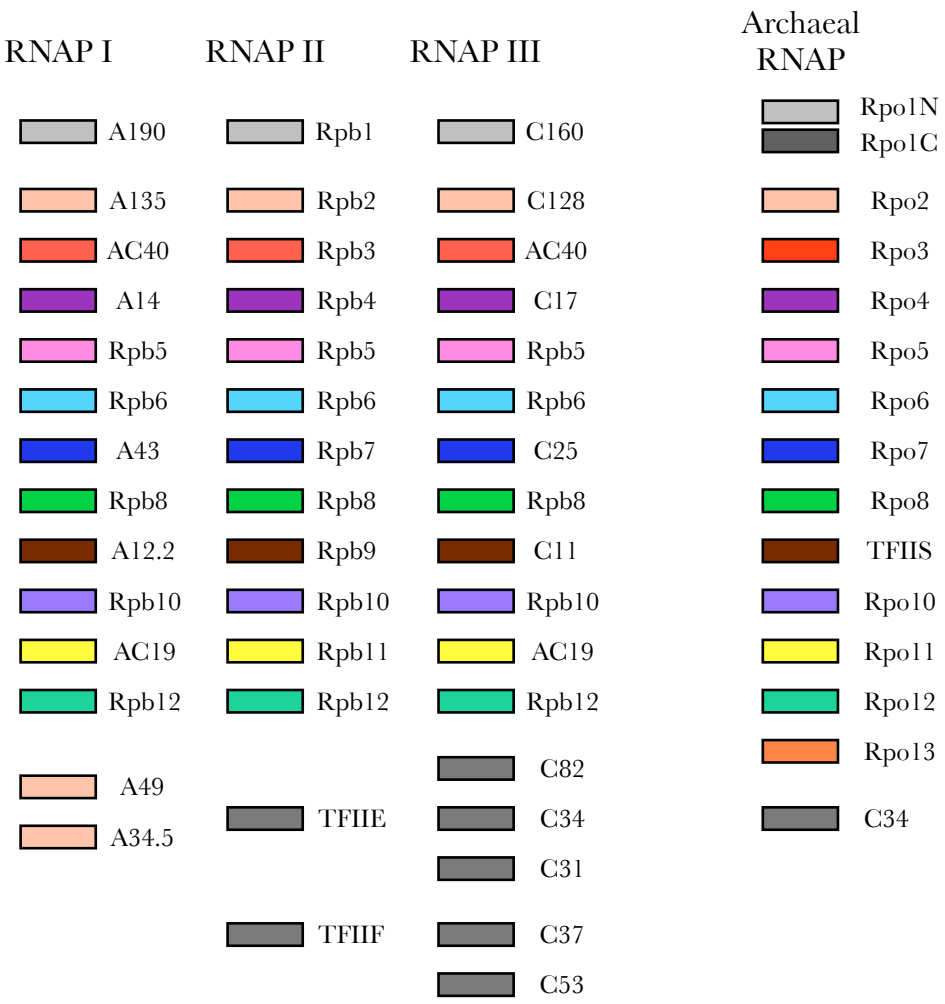


Figure 1.10: *Subunit composition in RNAP I, II and III. The Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12 subunits are shared between RNAP I, II and III. The crenarchaeal RNAP has 13 subunits homologous to the eukaryotic RNAPs, specifically RNAP-II, with the exception of Rpb9. Furthermore, the Rpo1 subunit is split into Rpo1N and Rpo1C. Moreover, the archaeal RNAP contains an archeaeal-specific extra subunit, denoted as Rpo13. In addition, an archaeal homologue of the RNAP-III-specific RPC34 subunit has been identified. The figure is recreated from [14].*

1.4.1 Archaeal RNAP's Overall Structure

The initial *S. solfataricus* RNAP crystal structures revealed that the relative positioning of the RNAP core and stalk are highly conserved between the archaeal RNAP and the three classes of eukaryotic RNAPs, in particular RNAP-II (Figure1.11) [79]. Conserved structures were found near the active centre and DNA binding channel [79]. Hence, the fundamentals behind the mechanism

of transcription are conserved between the archaeal and eukaryotic RNAPs. Furthermore, a striking structural conservation distributed throughout the entire structure in terms of overall architecture, subunit arrangement, composition and topology occurs between the two RNAPs with only slight differences. The largest Rpo1 subunit is split into two polypeptides in Archaea, which are encoded by separate genes in an operon, as opposed to the single eukaryotic Rpb1 subunit. Sequence alignments revealed that the archaeal Rpo1N and Rpo1C subunits correspond to the N-terminal two-thirds and the C-terminal heptad repeat of the Rpb1 subunit, respectively, and that the junction between Rpo1N and Rpo1C is positioned at the Rpb1 'foot' domain [79]. The archaeal 'foot' domain consists of four α-helices, which, despite a more complex architecture of the eukaryotic 'foot' domain, are conserved in the latter domain [79]. On the other hand, structural similarity between Rpb2 and Rpo2 is higher [79]. Upon solving of the crystal structure of the Rpo3/ Rpo11 complex, which forms a platform for assembly for the Rpo1 and Rpo2 subunits, it was found that the Rpo3 subunit contains an 3Fe-4S cluster [79], [107]. The 3Fe-4S cluster binding domain is unique to the *Sulfolobus* RNAP since the corresponding region in Rpb3 forms a loosely packed domain called the 'loop', and is thought to play a structural role, since its absence causes Rpo3 aggregation [79]. Also, four cysteine residues in the Rpo3 subunit form two disulfide bonds, while the same residues in Rpb3 chelate a zinc ion [79]. Finally, the major distinct structural differences between the two RNAPs are limited to the side of the face that faces downstream DNA in the transcription elongation complex (Figure1.11). Of all the subunits, Rpo5 differs the most in size to Rpb5, the latter being being composed of a jaw and assembly domain [46]. Only the assembly domain is present in Rpo5, since it lacks the first 130 residues of Rpb5. Thus, all the differences between the archaeal RNAP and RNAP-II can be classified as simple additions of RNAP-II-specific polypeptides to the archaeal RNAP, which include the addition of the eukaryotic-specific Rpb9 subunit and the Rpb1 C-terminal domain as well as the Rpb5 jaw domain, rather than changes to the core structure [79]. Hence, the archaeal

RNAP is not only a simplified version of its eukaryotic RNAP-II counterpart but also an excellent model system for understanding the molecular basis of eukaryotic transcription.

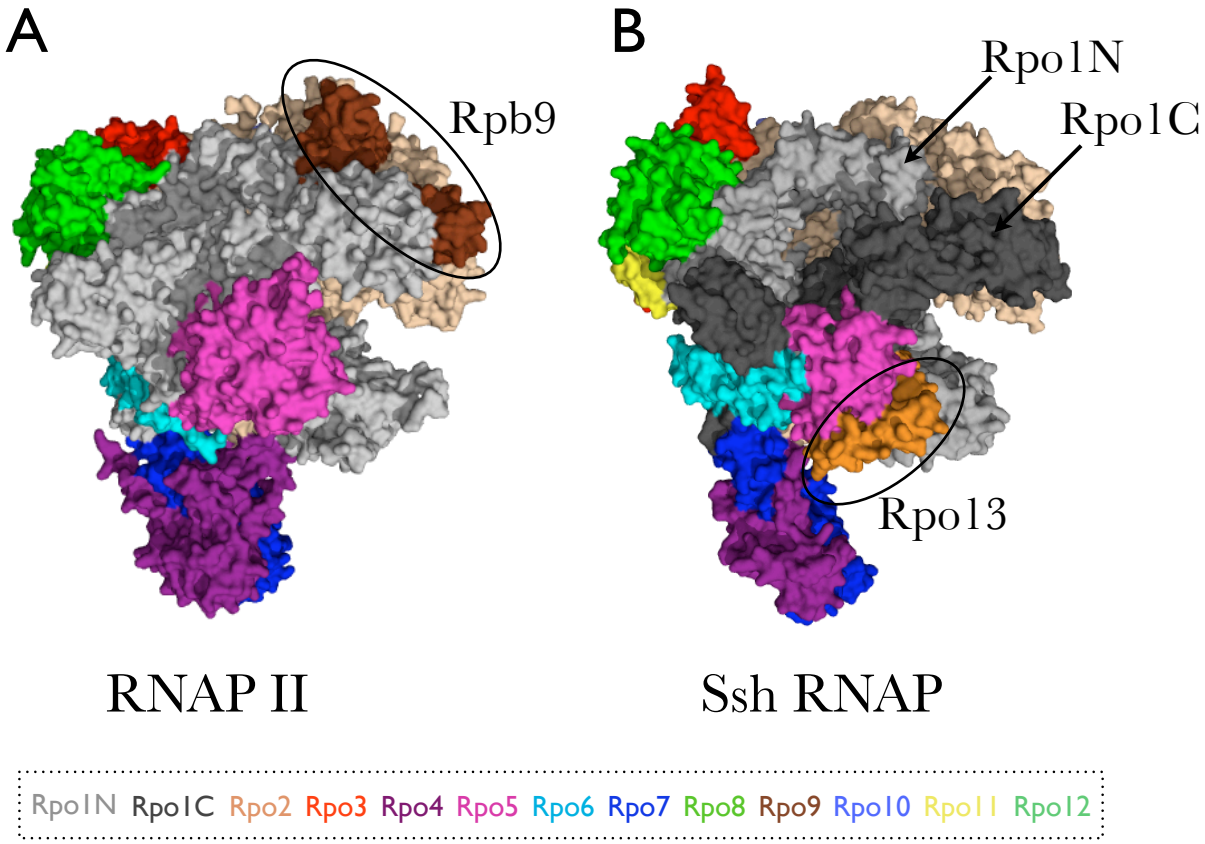


Figure 1.11: *Structural comparison of the eukaryotic RNAP-II with respect to that of the archaeal RNAP from S. shibatae (Ssh).* ***(A)*** *The eukaryotic RNAP-II contains 12 subunits, where Rpb9 is eukaryotic specific.* ***(B)*** *The archaeal Ssh RNAP is similar in its overall structure and subunit composition to RNAP-II. It contains 13 subunits which are homologous to RNAP-II, denoted as Rpo1-13. The Rpo1 subunit is split into Rpo1N and Rpo1C. Also the eukaryotic-specific Rpb9 subunit does not have an archaeal counterpart. Furthermore, a newly identified, archaeal-specific subunit denoted as Rpo13 has been identified, which occupies a position equivalent to the eukaryotic Rpb5 jaw domain. The figure is recreated from [96], where the PDB code for the RNAP-II crystal structure is 3FKI, while the PDB code for the archaeal RNAP crystal structure is 2WAQ.*

More importantly, collaboratory work performed by Dr. Nicola Abrescia and Prof. Stephen Bell have led to the resolution of the *S. shibatae* RNAP crystal structure, revealing new Rpo8 and Rpo13 subunits [96]. Indeed, it was initially believed that archaeal RNAPs from both the *Euryarchaeota* and *Crenarchaeota* did

not possess Rpb8 orthologs [170], [132]. Resolution of the *Pyrococcus furiosus* RNAP crystal structure from the *Euryarchaeota* kingdom revealed no density was present at the locations for the RNAP-II subunit Rpb8 [101]. However, divergent Rpb8 homologs were found in the *Crenarchaeota* ([102], [171]), where the Rpo8 subunit was initially defined ([104]), but its similarity to Rpb8 was not recognised since it shares the lowest sequence identity of all archaeal and eukaryotic subunits. Korkhin and coworkers subsequently confirmed that Rpo8 is a constitutive structural element of the *S. shibatae* RNAP [96]. Furthermore, a new density corresponding to a helix-turn-helix (HTH) motif with α1 and α2 helices and a third, positively charged C-terminal domain was found to fit in a groove created by the assembly domain of Rpo5 and the clamp-head domain of Rpo1N, which has a longer HTH motif than Rpb1's one [96]. This density corresponded to a previously reported RNAP subunit named component F [104], comprising 104 residues of which the 45-residue HTH motif constitutes an ordered fragment with flexible N- and C-termini [96]. The subunit was renamed as Rpo13. Thus, the Rpo13 subunit does not have an ortholog in RNAP-II and is unique to the archaeal RNAP, occupying a position equivalent to the Rpb5 jaw domain (Figure1.11). More specifically, sequence analysis showed that the *Rpo13* gene has orthologs in the orders *Sulfolobales* and *Desulfurococcales* of the *Crenarchaeota* as well as being present in the *Korarchaeota*, but not in the *Thaumarchaeota* and *Euryarchaeota* [96], [180]. Rpo13's presence was also confirmed also in the *S. solfataricus* RNAP structure obtained by Hirata and coworkers ([79]) upon comparison with the *S. shibatae* RNAP structure [96]. Finally, resolution of the first *S. shibatae* RNAP crystal structure in the presence of a bubble DNA substrate revealed that, while interactions of the RNAP with the DNA are equivalent to those observed in the RNAP-II-DNA crystal structure, further differences occur at the far end of the DNA-binding channel. Indeed, it was observed that the fork loop 2 as well as switches 1 and 2 are in close proximity to the template strand, with bending of the bridge helix and a disordered trigger loop similarly to the RNAP-II-DNA crystal structure [180]. On the other hand, stabilisation of the downstream DNA

occurs by the jaw lobe module's Rpo1 subunit, which is further in the clef due to the absence of an Rpb9's archaeal equivalent [180]. In addition, the DNA path deviates from the one of the RNAP-II-DNA complex ([180]) due to the absence of the Rpb5 jaw domain, which helps DNA docking and defines its trajectory. Also, no open to close movement of the clamp upon DNA binding occurs [180].

Furthermore, the Rpo4/7 stalk complex is stably associated with the archaeal RNAP ([68]), while the Rpb4/7 stalk is reversibly associated with the RNAP-II in yeast [130]. In addition to modulation of the clamp conformation and RNA binding both *in vitro* and *in vivo*, the Rpo4/7 stalk has been proposed to have the function of facilitating transcription bubble formation of the archaeal RNAP-promoter DNA complex ([127]) through interaction with the TFE factor ([133], [172]), which is monomeric and corresponds to the N-terminal half of the TFIIE-α subunit. Indeed, analysis of the Rpo4/Rpo7 stalk of the *S. shibatae* RNAP suggested that the C-terminal helix 6 (H6) of Rpo4 can move from its location, possibly contributing to the interaction with TFE [96]. Reconstitution experiments showed that addition of Rpo7 marginally increased promoter opening, while addition of both Rpo7 and TFE drastically increased promoter opening [127]. Also, Rpo7 catalyses open complex formation, while TFE stimulates bubble opening further at the upstream end of the bubble in the proximity of the transcription start site ([127], ([172]), which is mediated by interactions between the TFE WH domain and the non-template strand of the promoter DNA [119], [127]. Furthermore, TFE was found to stabilise open complexes in the absence of the Rpo4/7 stalk, indicating that TFE is able to interact with RNAP in the absence of the Rpo4/7 stalk, although TFE's effect is greatly enhanced by Rpo4's presence [127], [133], [172]. Thus, Rpo7 and TFE work synergistically during promoter opening. A similar situation parallels the Rpb4/7 stalk in RNAP-II, since it has been shown to interact with TFIIF, which interacts with promoter DNA [123].

1.4.2 Archaeal Transcription Factors

The basal archaeal transcription machinery is fundamentally related to the eukaryal one. In addition to the archaeal RNAP, which is related to RNAP-II, Archaea possess TBP and TFB, which are respectively related to the eukaryotic TBP and TFIIB/TFIIIB70. Specifically, archaeal TBPs lack any extensive N-terminal extension and display homology to the conserved C-terminal domain of eukaryotic TBP, with a unique and short acidic C-terminal tail in most Archaea [13]. Since no archaeal homologues of TAFs appear to exist, TBP is thought to exist as a homodimer in archaeal extracts [138]. *In vitro* transcription assays have shown that TBP in conjunction with TFB are sufficient for accurate and efficient transcription ([77], [138]), with TFE playing only a stimulatory role under conditions where promoter recognition by TBP is sub-optimal in the *Crenarchaeota* [12], [72]. This minimal subset of factors in conjunction with RNAP is able to catalyse *in vitro* transcription initiation even on topologically unfavourable templates [17]. This is in contrast with the multitude of eukaryotic factors involved in the RNAP-II basal apparatus, homologues of which are absent in Archaea except for TFIIH's subunits XPB and XPD, which however are involved in nucleotide excision repair [142], [147]. Thus, the opening of archaeal promoters is not ATP hydrolysis-driven. Instead, it was proposed that thermal energy is required to facilitate DNA melting, since it was found that localised melting on all templates occurs at higher temperatures similar to the ones needed for organism growth [17]. On the other hand, the reduced thermodynamic input at lower temperatures was insufficient to assist the RNAP in overcoming the resistance to melting in the overwound, positively supercoiled DNA [17].

Co-crystallisation of TBP and a C-terminal core domain of TFB on an archaeal promoter DNA revealed that TBP and TFB recognise respectively TATA and BRE promoter elements ([97], [110], [139], [18]), which are conserved in archaeal promoters similarly to eukaryotic promoters. Specifically, while the overall fold of TBP and TFB is similar to the eukaryotic counterparts, the archaeal TBP-DNA interface was

found to be more symmetrical due to more perfect TBP and TFB repeats, where upon analysis of the crystal structure it was initially suggested that the orientation of the pre-initiation complex assembly on the promoter is inverted with respect to that observed in the equivalent crystal structures of eukaryotic systems [97]. However, subsequent studies showed that the TFB's interaction with the BRE element is the principal determinant of transcriptional polarity in Archaea, where the solution orientation of the TBP and TFB complex on the DNA was found to be the same as in the eukaryal complex due to the presence of the BRE element previously missing in the crystal structure [18], [110]. Therefore, TBP and TFB cooperatively bind to DNA [138], [15], [146], [73]. Furthermore, the N-terminal zinc-ribbon domain of TFB is involved in RNAP recruitment through the Rpo6 subunit ([15], [116], [172]), where, upon establishment of a stable pre-initiation complex, promoter clearance and productive transcription is facilitated by a conserved TFB motif located between the N-terminal zinc-ribbon and C-terminal core domain [15], [172]. A highly conserved Glu46 residue was found to affect TSS selection once mutated ([15]), similarly to the Glu62 residue in TFIIB [98]. Furthermore, TFB's action is complemented by the one of TFE, which shows that both of these factors act synergistically during transcription initiation [172].

More importantly, while *in vitro* transcription reconstitution assays with purified RNAP and recombinant TBP and TFB showed that the same transcription levels from various promoters can be obtained as when directed by crude *S. shibatae* cell extracts ([138]), one exception to this general finding was the structural 16S/23S rRNA promoter, where different transcription levels were observed [138]. This suggests the existence of additional regulatory factors, which may be involved in the transcription of non-coding genes in addition to the basal archaeal transcription machinery. Indeed, recent analysis of the *Thaumarchaeota* and *Crenarchaeota* archaeal genomes has identified a homologue of the eukaryotic RNAP-III-specific RPC34 subunit [22]. The archaeal RPC34 was found to have a WH domain as well as a zinc-finger domain shared with most eukaryotic RPC34 sequences

[22]. Analysis of the genomic context of the archaeal RPC34 revealed the co-localisation of this gene with factors involved in the transcription, modification and processing of structural rRNAs and tRNAs, such as a tRNA pyrophosphorylase and a paralogue of TFB referred to as TFB-2 [22]. Indeed, multiple archaeal TBPs and TFBs have been reported to be functional [10]. TFB-2 may therefore interact with the archaeal RPC34 for the transcription of structural promoters similarly to the eukaryotic interaction between RPC34 and TFIIIB70 [29], [91], [173]. Thus, RPC34 may be involved in the transcription of non-coding rRNA and tRNA genes. More importantly, it may associate with the single archaeal RNAP, modulating the specificity of the archaeal RNAP and generating a precursor of the eukaryotic RNAP-III. The split of RNAP into the various forms of RNAPs dedicated to the transcription of specific protein-coding and structural genes may have thus already occurred in Archaea. While Qureshi and coworkers showed that transcription from the 16S/23S rRNA promoter could be successfully initiated in the absence of RPC34 indicating that there is no strict RPC34 requirement for promoter recognition and RNAP recruitment, the different transcription levels obtained by the reconstituted system and crude cell extract suggest that additional factors such as RPC34 might involved in the regulation of the promoter [138]. The WH domain might mediate protein-DNA interactions, considering that the eukaryotic RPC34 was cross-linked to DNA in transcription initiation complexes [11]. Alternatively, RPC34 might simply help the recruitment of the archaeal RNAP through interaction with its clamp domain similarly to the RNAP-III system ([136]), increasing transcription levels, or may have no role in transcription.

Chapter 2

Experimental Aims

1. The newly identified Rpo13 subunit occupies a position in the archaeal RNAP which, in RNAP-II, is occupied by the eukaryotic-specific Rpb5 jaw domain which is involved in downstream DNA binding. It also has a third positively charged tail. Therefore, it has been proposed that Rpo13 might interact with DNA, where, at initiation, the C-terminal third tail could provide a lock point, against which the main RNAP body can push or twist the DNA, while $\alpha 1$ and $\alpha 2$ helices act as constitutive anchors onto the RNAP [96]. During elongation, these locking interactions would be overcome by translocation forces but may still affect DNA duplex stability [96]. Thus, Rpo13 may perform some of the roles attributed to the eukaryotic-specific initiation factors missing in Archaea. The various aims for the Rpo13 study are:

   - **To clone, express and purify for the first time the newly identified archaeal-specific Rpo13 subunit.**

   - **To raise an antibody against the newly purified Rpo13.**

   - **To test for an association between Rpo13 and DNA, to assess whether the third positively charged tail is the major**

determinant of DNA binding, and whether Rpo13 associates with the major or minor groove of DNA or in a way to introduce DNA torsional stress to generate supercoiled DNA.

- **To test whether Rpo13 is a *bona fide* subunit of the archaeal RNAP, or whether it temporarily associates with the RNAP, in similarity to a transcription factor as opposed to a true subunit.**

- **To study the function of Rpo13 in the context of transcription.**

2. The newly identified archaeal RPC34 homologue was found to be co-localised from a genomic point of view with factors involved in the transcription, modification and processing of structural rRNAs and tRNAs, such as a tRNA pyrophosphorylase and a paralogue of TFB referred to as TFB-2 [22]. RPC34 may therefore associate with TFB-2 in analogy to TFIIIB70, and could be involved in the recruitment of the archaeal RNAP in similarity to the recruitment of RNAP-III by the RPC34-TFIIIB70 complex [29], [91], [173]. The archaeal RNAP would be directed by the RPC34-TFB-2 complex to the transcription of non-coding genes, where RPC34 would have a regulatory role since it would increase transcription levels of this subset of genes. RPC34's association with the single archaeal RNAP would therefore define a precursor of the eukaryotic RNAP-III, where RPC34 would modulate the specificity of the archaeal RNAP. The various aims for the RPC34 study are:

   - **To clone, express and purify for the first time the newly identified RPC34 archaeal homologue.**

   - **To raise an antibody against the newly purified RPC34.**

- **To test for an *in vitro* and *in vivo* interaction between RPC34 and the single archaeal RNAP.**

- **To test for an *in vitro* interaction between RPC34 and TFB-2.**

- **To co-express recombinant RPC34 in conjunction with recombinant TFB-2 to test whether a soluble complex is obtained.**

- **To test whether RPC34 is cell cycle regulated in similarity to TFB-2, in a way to associate with the archaeal RNAP only at a certain cell cycle stage.**

3. Rpo13 was found to be an integral subunit of the archaeal RNAP upon crystallisation of the latter [96]. Nonetheless, it remains to be tested whether Rpo13 is a *bona fide* subunit of the archaeal RNAP *in vivo* in its elongation stage, or whether it temporarily associates with the archaeal RNAP in a transcription-factor like mode. The various aims for the study of Rpo13's genomic location are:

   - **To use the antibody raised against Rpo13 in conjunction with the one previously raised against Rpo2 in order to perform a chromatin-immunoprecipitation (ChIP)-sequencing experiment which would yield insights into the genomic distribution of Rpo2 versus the one of Rpo13.**

   - **To perform a RNA-sequencing experiment, which, in conjunction with the previous ChIP-sequencing experiments, would allow for the comparison of RNAP distribution versus**

transcript expression and abundance.

Chapter 3

Materials and Methods

3.1 Cloning

All primers were bought from Eurogentec S.A and are described in (Table3.1) and (Table3.2). Open reading frames were amplified by polymerase chain reaction (PCR) and cloned into expression plasmid vectors via restriction enzyme digestion (NEB) as described in (Table3.1) and (Table3.2) with a subsequent ligation reaction catalysed by T4 DNA ligase (NEB). *E.coli* XL-10 competent bacterial cells (Tetr Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F proAB lacIqZΔM15 Tn10 (Tetr) Amy Cam^r] were transformed with the ligation reactions. Specifically, *E.coli* cells were incubated with the ligation reactions for 30 minutes at 4 °C on ice, followed by a heat-pulse in a 42 °C water bath for 30 seconds and a subsequent incubation on ice for 2 minutes. One ml of pre-heated luria-bertani (LB) medium was added to the ligation/bacterial cell sample mix, where the sample was incubated at 37 °C for one hour with shaking at 225-250 rpm. The transformation mixture was then plated on LB agar plates containing the appropriate antibiotic (Table3.1) and (Table3.2) for on-night incubation at 37°C. Colony PCR was performed for the screening of bacterial colonies containing the plasmid with the correct insert, where positives where grown overnight, followed by plasmid isolation and sequencing (Geneservice UK, Oxford).

Gene	Saci_number	Expression Plasmid	Primers	R.E. used	Antibiotic
Rpo13 *S. acidocaldarius*	Saci_0816	pET30a	**F** GGG**CATATG**TCTGAGGACGATTCTAAAAAG **R** GGG**GAATTCTCA**TTCTTCTTCCTCATCAGATTTC	NdeI EcoRI	Kanamycin
Rpo13 *S. acidocaldarius* (His-tag)	Saci_0816	pET30a	**F** GGG**CATATG**TCTGAGGACGATTCTAAAAAG **R** GGG**GAATTC**TTCTTCTTCCTCATCAGATTTCTTAG	NdeI EcoRI	Kanamycin
Rpo13 *S. shibatae* (No His-tag)		pET30a	**F** GATTCCC**CATATG**GTATCTGGTATGTCAACAG **R** GAATTTC**CTCGAGTTA**ACCCTCTACACTCTTTTC	NdeI XhoI	Kanamycin
Rpo13 *S. shibatae* (His-tag)		pET30a	**F** GATTCCC**CATATG**GTATCTGGTATGTCAACAG **R** GAATTTC**CTCGAG**ACCCTCTACACTCTTTTC	NdeI XhoI	Kanamycin
RPC34 *S. acidocaldarius*	Saci_1342	pET33b	**F** GAATTC**CCATGG**GGAAAGAGAATATATTATTTG **R** GAATTC**CTCGAGTTA**TCCAGCAGAAACACTAGTT	NcoI XhoI	Kanamycin
RPC34-GST *S. acidocaldarius*	Saci_1342	pET42a	**F** GAATTC**CCATGG**GGAAAGAGAATATATTATTTG **R** GAATTC**CTCGAGTTA**TCCAGCAGAAACACTAGTT	NcoI XhoI	Kanamycin
RPC34-His *S. acidocaldarius*	Saci_1342	pET33b	**F** GAATTC**CCATGG**GGAAAGAGAATATATTATTTG **R** CCC**CTCGAG**TCCAGCAGAAACACTAGTTAAG	NcoI XhoI	Kanamycin
TFB-2 Entire *S. acidocaldarius*	Saci_1341	pET33b	**F** GAATT**CCATGG**AATGTAAAATCTGTGGCAGTGAG **R** GGAATTC**CTCGAG**AAGCTTTACTTCTATGTCAAAAGCCTTT	NcoI XhoI	Kanamycin
TFB-2 Core *S. acidocaldarius*	Saci_1341	pET33b	**F** GAATTC**CCATGG**AAATAGGATATACTAAAATCAAAGACCGAATAAA **R** GGAATTC**CTCGAG**AAGCTTTACTTCTATGTCAAAAGCCTTT	NcoI XhoI	Kanamycin
TFB-2 *S. acidocaldarius*	Saci_1341	pET-Duet-1	**F** GAATT**CCATGG**AATGTAAAATCTGTGGCAGTGAG **R** GGAATTC**CTCGAG**AAGCTTTACTTCTATGTCAAAAGCCTTT	NcoI XhoI	Ampicillin
RPC34 *S. acidocaldarius*	Saci_1342	pRSF-Duet-1	**F** GAATTC**CCATGG**GGAAAGAGAATATATTATTTG **R** GAATTC**CTCGAGTTA**TCCAGCAGAAACACTAGTT	NcoI XhoI	Kanamycin

Table 3.1: *A table showing all the genes cloned into their corresponding plasmid expression vectors with the described primers. All the primers were used to amplify the genes with PCR reactions, where the PCR amplicon was subsequently digested with the appropriate pair of restriction enzymes and ligated into the vectors. Selection of positive E.coli transformants was performed with the correct antibiotic, which selects for plasmid-containing E.coli cells.*

3.2 Recombinant Protein Purification

E. coli Rosetta bacterial cells [(F^- *ompT* $hsdS_B(r_B{}^-m_B{}^-)$ *gal dcm* (DE3) pLysSpRARE2[3] (Cam^R)] transformed with the engineered plasmids were grown in LB medium at 37 °C to OD_{600} = 0.6 and induced with 1 mM IPTG for 4 h.

Cultures expressing Rpo13 protein were centrifuged, where cell pellets were resuspended in buffer containing 10 mM of Tris-HCl (pH 8) and 50 mM of NaCl. After cell disruption with a French press and centrifugation at 17 000 rpm with a JA-25-50 Beckman rotor, the supernatants were incubated at 70 °C for 20 min prior to protein purification. Both the *S. shibatae* and *S. acidocaldarius* Rpo13 proteins were purified by S-sepharose cation exchange chromatography (Figure3.1), where only the initial fractions were collected for *S. acidocaldarius* Rpo13 due to a

Gene	Saci_number	Yeast-two hybrid Plasmid	Primers	R.E. used	Antibiotic
RPC34 *S. acidocaldarius*	Saci_1342	pGBKT7	**F** GAATTC**CCATGG**GGAAAGAGAATATATTATTTG **R** CCC**GTCGACTTA**TCCAGCAGAAACACTAGTT	NcoI SalI	Kanamycin
RPC34 *S. acidocaldarius*	Saci_1342	pGADT7	**F** GAATTC**CCATGG**GGAAAGAGAATATATTATTTG **R** GAATTC**CTCGAGTTA**TCCAGCAGAAACACTAGTT	NcoI XhoI	Ampicillin
Rpo13 *S. acidocaldarius*	Saci_0816	pGADT7	**F** GGG**CATATG**TCTGAGGACGATTCTAAAAAG **R** GGG**GAATTCTCA**TTCTTCTTCCTCATCAGATTTC	NdeI EcoRI	Ampicillin
TFB-2 *S. acidocaldarius*	Saci_1341	pGBKT7	**F** GGG**GAATTC**ATGAAATGTAAAATCTGTGG **R** CCC**GTCGACTCA**AAGCTTTACTTCTATGTC	EcoRI SalI	Kanamycin
TFB-2 *S. acidocaldarius*	Saci_1341	pGADT7	**F** GGG**GAATTC**ATGAAATGTAAAATCTGTGG **R** CCC**CTCGAGTCA**AAGCTTTACTTCTATGTC	EcoRI XhoI	Ampicillin
TFB *S. acidocaldarius*	Saci_0866	pGBKT7	**F** GGG**CATATG**GGTGGTGAAATGGTAG **R** CCC**GTCGACTTA**TTGACTTGGAATCTG	NdeI SalI	Kanamycin
TFB *S. acidocaldarius*	Saci_0866	pGADT7	**F** GGG**CATATG**GGTGGTGAAATGGTAG **R** CCC**CTCGAGTTA**TTGACTTGGAATCTG	NdeI XhoI	Ampicillin

Table 3.2: *A table showing all the genes cloned into their corresponding yeast-two hybrid plasmid vectors with the described primers. All the primers were used to amplify the genes with PCR reactions, where the PCR amplicon was subsequently digested with the appropriate pair of restriction enzymes and ligated into the vectors. Selection of positive E.coli transformants was performed with the correct antibiotic, which selects for plasmid-containing E.coli cells.*

contaminant *E.coli* protein co-eluting with Rpo13 at later fractions. Specifically, a 5 ml S-sepharose HiTrap SP HP (GE Healthcare Life Sciences) column was used, where the fluid flow rate was 2 ml/min and the NaCl elution gradient performed over a 50 ml volume starting from 0% to 100% of 1 M NaCl with collection of 1.5 ml elution fractions, where elution of Rpo13 occurs at 60%-70% of 1 M NaCl (Figure3.1). The identity of Rpo13 was confirmed by mass spectrometry (Oxford University Central Proteomics Service), with a tryptic digest followed by MALDI mass spectrometry and Mascot Database search.

Cultures expressing mutant Rpo13 proteins lacking either the negatively charged N-terminal protein region of Rpo13 (ΔN Rpo13), or the the third, positively charged C-terminal tail (ΔC Rpo13), or both of these protein regions (core Rpo13) were centrifuged. ΔC Rpo13 and core Rpo13 cell pellets were resuspended in buffer containing 10 mM of Tris-HCl (pH 8) and 300 mM of NaCl with subsequent cell disruption with a French press and centrifugation at 17 000 rpm with a JA-25-50

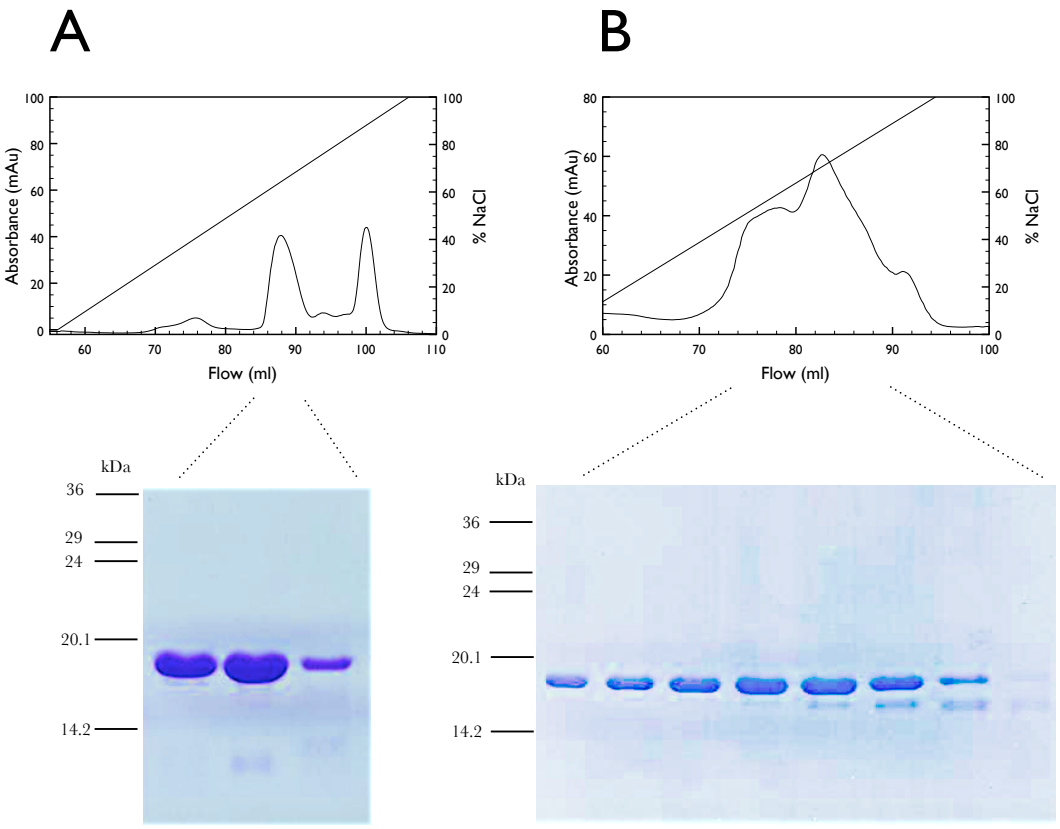


Figure 3.1: *Rpo13's purification with affinity S-sepharose cation exchange chromatography.* ***(A)*** *Elution profile for S. shibatae Rpo13 protein, where it is found that around 60%-70% of 1 M NaCl is required for its elution. Rpo13 is found to migrate at a higher molecular weight of around 20 kDa than its predicted 12 kDa.* ***(B)*** *Elution profile for S. acidocaldarius Rpo13 protein, where it is found that around 60%-70% of 1 M NaCl is required for its elution. Only the initial fractions were pooled together for sample storage of S. acidocaldarius Rpo13 protein due to a contaminant E.coli protein co-eluting with Rpo13 at later fractions. Rpo13 is found to migrate at a higher molecular weight of around 20 kDa than its predicted 12 kDa.*

Beckman rotor. All three Rpo13 protein mutants were purified by gravity-flow Ni^{2+}-NTA affinity chromatography with a subsequent gel filtration step. Specifically, the Rpo13 mutants were eluted from the Ni^{2+}-NTA beads with 500 mM imidazole, after one initial wash with buffer containing only 10 mM of Tris-HCl (pH 8) and either 300 mM of NaCl or 900 mM of NaCl and a second wash with the equivalent buffers containing 20 mM imidazole. A HiLoad 26/60 Superdex 75 pg, 320 ml, gel filtration column (GE Healthcare Life Sciences) equilibrated with buffer containing 10 mM of Tris-HCl (pH 8) and either 300 mM of NaCl or 900 mM of NaCl was subsequently used, where the fluid flow rate was 1.5 ml/min with collection of 2 ml

elution fractions. The ΔC Rpo13 mutant is found to elute at an elution volume of 180 ml, while the ΔN Rpo13 mutant is found to elute at an elution volume of 200 ml, and finally the core Rpo13 mutant is found to elute at an elution volume of 240 ml. The His-tag of the ΔC Rpo13 mutant was cleaved with the HRV 3C protease (Novagen). The purity of the various Rpo13 mutants is shown in (Figure3.2).

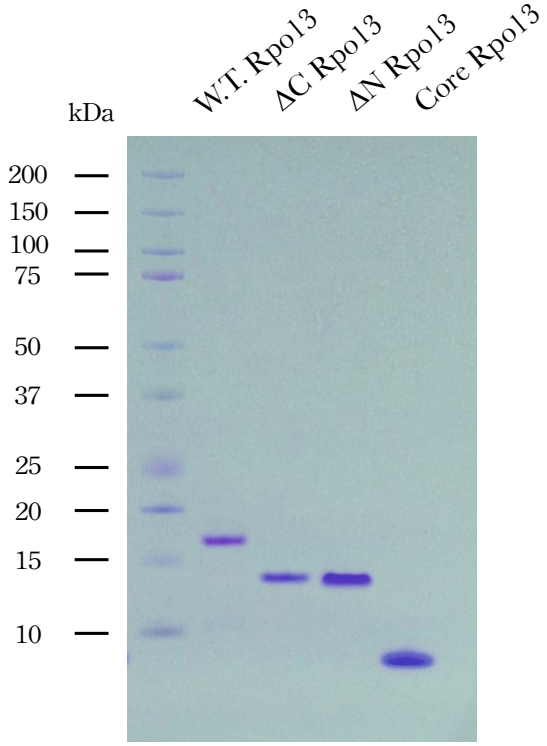


Figure 3.2: *SDS-PAGE gel subsequently stained with Coomassie Brilliant Blue showing purified Rpo13 mutants. These Rpo13 mutants lack either the negatively charged N-terminal region of Rpo13 (ΔN Rpo13), or the the third, positively charged C-terminal tail (ΔC Rpo13), or both of these regions (core Rpo13). The mutants are found to migrate according to their molecular weights of 9.4 kDa for ΔC Rpo13, 9.7 kDa for ΔN Rpo13 and 6.7 kDa for core Rpo13.*

Cultures expressing recombinant, His-tagged RPC34 were centrifuged, where cell pellets were resuspended in buffer containing 10 mM of Tris-HCl (pH 8), 300 mM of NaCl and 8 M urea with subsequent cell disruption with a French press and centrifugation at 17 000 rpm with a JA-25-50 Beckman rotor. His-tagged RPC34 was purified by gravity-flow Ni^{2+}-NTA affinity chromatography under denaturing conditions with elution from the Ni^{2+}-NTA beads with 500 mM imidazole, after

one initial wash with buffer containing only 10 mM of Tris-HCl (pH 8), 300 mM of NaCl and 8 M urea and a second wash with the equivalent buffer containing 20 mM imidazole.

Cultures expressing recombinant, His-tagged TFB-2 in either its entire or core form were centrifuged, where cell pellets were resuspended in buffer containing 10 mM of Tris-HCl (pH 8), 300 mM of NaCl and 4 M guanidinium hydrochloride with subsequent cell disruption with a French press and centrifugation at 17 000 rpm with a JA-25-50 Beckman rotor. Both His-tagged entire and core TFB-2 were purified by gravity-flow Ni^{2+}-NTA affinity chromatography under denaturing conditions with elution from the Ni^{2+}-NTA beads with 500 mM imidazole, after one initial wash with buffer containing only 10 mM of Tris-HCl (pH 8), 300 mM of NaCl and 4 M guanidinium hydrochloride and a second wash with the equivalent buffer containing 20 mM imidazole.

Finally, TBP and TFB proteins were purified as previously described [15]. TBP was purified by S-sepharose cation exchange with a 5 ml S-sepharose HiTrap SP HP (GE Healthcare Life Sciences) column, where the fluid flow rate was 2 ml/min and the NaCl elution gradient performed over a 100 ml volume starting from 0% to 100% of 1 M NaCl with collection of 1.5 ml elution fractions, with a subsequent gel filtration step with a HiLoad 26/60 Superdex 75 pg, 320 ml, gel filtration column (GE Healthcare Life Sciences), where the fluid flow rate was 1.5 ml/min with collection of 2 ml elution fractions. TFB was purified by gravity-flow Ni^{2+}-NTA affinity chromatography with a subsequent gel filtration step with a HiLoad 26/60 Superdex 75 pg, 320 ml, gel filtration column (GE Healthcare Life Sciences), where the fluid flow rate was 1.5 ml/min with collection of 2 ml elution fractions.

3.3 Antibody Raising against Recombinant Rpo13 and RPC34

Polyclonal antibodies against recombinant, native Rpo13 and recombinant, unfolded RPC34 in the presence of 4 M urea were raised upon protein injection into rabbits (Eurogentec), where an immune response was generated upon four injections of 200 μg of antigen per injection. A Hi-Trap NHS-activated HP column (GE Healthcare) was used to affinity-purify antibodies against Rpo13 from the serum, where Rpo13 protein was coupled to the column with the serum left to circulate for 2 hours through the column. Buffer containing 100 mM glycine (pH 2.6) in 100 mM Tris was used to elute bound antibodies, where 100 μl of 1 M untitrated Tris was immediately added in the 1 ml elution fractions to neutralise the pH.

3.4 EMSA Analysis

Electrophoretic mobility shift assays (EMSAs) involving TBP, TFB and the TBP-TFB fusion were performed with ^{32}P radiolabeled dsDNA T6 promoter sequence containing a TATA-like A box ([139]), while EMSAs involving wild-type Rpo13 as well as Rpo13 mutants were performed with the non-specific DNA sequence of 40 bp (5' ATATTTACCTTAAGTTCTAACGTGGAAACAAAGGGGTTTT 3) in either ssDNA form or annealed to its complementary sequence for dsDNA formation. Protein incubation with DNA was performed in 50 mM Tris pH = 8, 25 mM $MgCl_2$ and 75 mM KCl. PolydGdC was added only in EMSAs involving TBP, TFB and covalently linked TBP-TFB. Incubations were performed for 15 min at 48 °C. Protein-DNA complexes were run on native Tris-glycine gels at 170 V for 2 hours. Competition assays with methyl green and netropsin were adapted from [155].

3.5 EMSA Analysis coupled to Western Blot Analysis

An EMSA procedure subsequently coupled to western blot analysis was performed [50]. In this procedure, the EMSA gel is coupled with a nylon membrane and a nitrocellulose membrane for transfer of protein-DNA complexes by wet electroblotting, where the nylon membrane retains DNA while the nitrocellulose membrane retains protein. A standard western blot procedure was subsequently performed on the nitrocellulose membrane. The nitrocellulose membrane was blocked with 5% dry milk in TBS-T (TBS and 0.1% Tween 20) for 1 h and then incubated for 1 h with anti-Rpo13 rabbit antibody diluted 1000-fold in TBS-T with 5% milk. Three consecutive washes of each 10 minutes were performed with TBS-T with 5% milk, followed by incubation with an anti-rabbit IgG horseradish peroxidase (HRP) conjugate antibody (Fisher Scientific) diluted 10000-fold in TBS-T with 5% milk for 45 min and further three consecutive washes of each 10 minutes with TBS-T with 5% milk with a final wash with TBS-T without milk. ECL reagents (Amersham) were applied on the nitrocellulose membrane in order to develop western blots.

3.6 Supercoiling Assays

Supercoiling assays based on a nick-closure plasmid system were adapted from [117] and [70]. The pBR322 plasmid containing a single nick per circular molecule was prepared by digestion with the nicking enzyme Nb.Bpu10I (NEB). The nicked plasmid was subsequently incubated with Rpo13 protein at 48 °C for 15 min, and ligated with the T4 DNA ligase (NEB). Samples were deproteinised with proteinase K, phenol-chloroform extracted followed by DNA precipitation in the presence of ethanol and ammonium acetate, and analysed by agarose gel electrophoresis in either the absence of ethidium bromide for subsequent, post-run staining with SYBR Green I, or in the presence of ethidium bromide with a subsequent Southern blot procedure. A 0.5 μg of pBR322 plasmid was used per reaction.

For the Southern blot procedure, the agarose gel was soaked firstly in 0.25 M HCl, then in 0.5 M NaOH and 10X SSC buffer. Upwards capillary transfer of the DNA onto the nylon membrane was achieved in 0.5 M NaOH and 10X SSC buffer, with subsequent cross-linking of the DNA to the membrane with a UV cross-linker. The membrane was then pre-rinsed with the hybridisation Church-Gilbert buffer (7% SDS, 0.5 M $NaPO_4$ buffer pH = 7.2, 1 mM EDTA) at 65 °C for 6 hours, where a random-prime labelled reaction probe sample, obtained with 5 μl octadeoxyribonucleotides (400 mg/ml) and a 6 μl dNTP mixture (2 μl of dCTP, dTTP, and dGTP, each of a concentration of 0.5 mM) all from the NEB random-priming reaction kit as well as 5 μl of [α-^{32}P]-dATP (3000 ci/mmol, 50 μCi) in the presence of DNA polymerase and the pBR322 plasmid, was added. The probe was left to hybridise overnight. Washes were performed firstly with 2X SSC, 0.1% SDS buffer twice and then with 0.5X SSC, 0.1% SDS buffer for three times. The membrane was finally exposed to a phosphor screen.

3.7 *In Vitro* Transcription Assays with TBP, TFB and the TBP-TFB Fusion

In vitro transcription assays were performed as previously described [83], [137], [17], [18]. Briefly, reactions containing 200 ng of RNAP and transcription factor components were assembled with the rNTP mix and a final buffer concentration of 50 mM Tris pH = 8, 25 mM $MgCl_2$ and 75 mM KCl as well as 100 ng of pT6 template plasmid DNA. Reactions of a 50 μl final volume were incubated at 70 °C for 20 min. Reactions were stopped with buffer containing 10 mM Tris pH = 8, 10 mM EDTA, 750 mM NaCl, 1% SDS, and that also included 1 μl of ^{32}P radiolabeled T7 sequencing primer as well as 10 μg of glycogen. Reactions were phenol-chloroform extracted, with subsequent RNA precipitation in ethanol, where RNA was reverse transcribed into DNA with AMV reverse transcriptase (Roche). DNA samples were subsequently analysed on a denaturing 8 M urea/acrylamide sequencing gel.

3.8 *In Vitro* Transcription Assays with Rpo13

In vitro transcription assays were adapted from [122]. Briefly, reactions containing 200 ng of RNAP and transcription factor components as well as DNA substrates were assembled with a 1.25 μl rNTP mixture of 2 mM and lacking rUTP, 0.06 μl of rUTP (2mM) and 0.5 μl of [α-^{32}P]-UTP (3000 ci/mmol) in a final buffer concentration of 50 mM Tris pH = 8, 25 mM $MgCl_2$ and 75 mM KCl. Reactions of a 12.5 μl final volume were incubated at 70 °C for 20 min, and stopped with buffer containing 10 mM Tris pH = 8, 10 mM EDTA, 750 mM NaCl and 1% SDS. Reactions were phenol-chloroform extracted followed by RNA precipitation in the presence of ethanol and glycogen. RNA samples were subsequently analysed on a denaturing 8 M urea/acrylamide sequencing gel.

3.9 Attempts at removing Rpo13 from the RNAP

The RNAP sample obtained as a gift from Dr Nicola Abrescia was passed through a MonoQ column as described in [104]. Specifically, a mono Q HR 16/10 (20 ml) anion exchange column (GE Healthcare Life Sciences) was used, equilibrated in buffer containing 10 mM ammonium acetate and 20 mM Tris-HCl (pH 7.6), where the fluid flow rate was 0.5 ml/min. An ammonium acetate gradient was subsequently applied, increasing by 25% of 1 M ammonium acetate for the first 10 ml, then by 50% of 1 M ammonium acetate for 30 ml, and again by 25% of 1 M ammonium acetate for the last 10 ml, with collection of 2 ml elution fractions. Furthermore, incubation of 40 μg His-tagged Rpo13 with 0.4 μg of RNAP for exchange of endogenous Rpo13 with recombinant Rpo13 was performed in 2X TBS buffer containing either 0.5 M NaCl in conjunction with 1% Triton X-100 or 1 M NaCl. The samples were subsequently passed onto Ni^{2+}-NTA beads, where a western blot with the antibody against Rpo2 was subsequently performed (1:1000 dilution) to test whether His-tagged Rpo13 bound to the Ni^{2+}-NTA beads also pulls down the RNAP.

3.10 Cell Extract Preparation

A *S. acidocaldarius* culture was grown in Brock's *Sulfolobus* medium ([26]) until OD_{600} = 0.4. Specifically, 1 X Brock's medium was made from a 1000 X stock solution ($CaCl_2$ X $2H_2O$, 70 g/l), a 200 X stock solution (KH_2PO_4, 56 g/l, 10 nM $MnCl_2$, 120 nM $Na_2B_4O_7$, 640 nM $ZnSO_4$, 0.5 nM $CuCl_2$, 0.1 nM $NaMoO_4$, 0.1 nM $VOSO_4$, 0.03 nM $CoSO_4$, 0.09 nM $NiSO_4$, 3 ml 50% H_2SO_4), and a 100 X stock solution ($(NH_4)_2SO_4$, 130 g/l, $MgSO_4$ X $7H_2O$, 25 g/l, $FeCl_2$, 2 g/l, 3 ml 50% H_2SO_4), where 2 g of tryptone was added and the pH of the solution was adjusted to 3.2 with 50% H_2SO_4. The culture was subsequently centrifuged, where the cell pellet was resuspended in 1 X TBS buffer. Triton X-100 was added to 0.1% for cell lysis, where a subsequent centrifugation step at maximum speed with the JA20 rotor was performed at 4 °C for collection of the supernatant. The supernatant was subsequently passed through a 0.45 μM filter syringe.

3.11 GST-Pulldown Assay

A 10 μl bead volume containing either 10 μg of GST or RPC34-GST fusion protein was incubated with 10 μl of *S. acidocaldarius* cell extract (3 mg/ml). Incubations were performed for 30 min at 30 °C in 100 μl of 10 mM Tris pH = 8, 150 mM NaCl, 0.1% Tween-20, 5 mM EDTA. Beads were washed three times with 500 μl of the incubation buffer and resuspended in 100 μl SDS-PAGE loading buffer, where 10 μl of samples were electrophoresed. A western blot was subsequently performed with the antibody against Rpo2 (1:1000 dilution) to test whether RPC34-GST fusion protein bound to glutathione agarose beads pulls down the RNAP.

3.12 Yeast-two Hybrid Screen

A matchmaker Gal4 two hybrid system featuring the yeast strain AH109 and the *his3* gene was used (Clontech). A yeast culture was grown in YEP medium until OD_{600} =

0.4 at 30 °C, where 1 ml of culture was taken per co-transformation procedure, with a subsequent centrifugation step and wash with 1 ml of sterile water. The cell pellet was then resuspended in 50 μl of 1 X TE/1 X LiAc, with 10 X TE being 100 mM Tris (pH 8), 10 mM EDTA (pH 8) and 10 X LiAc being 1 M LiAc (pH 7.5), where 2-5 μl of plasmid DNA (0.1-1 μg) was added for each plasmid to be co-transformed. Subsequently, 25 μl of 2 mg/ml ssDNA was added in conjunction with 300 μl of PEG solution (40% PEG 4000, 1 X TE, 1 X LiAc), followed by vortexing. Samples were incubated at 30 °C for 2 hours with a subsequent heat-shock in a water bath at 42 °C for 15 minutes. Samples were finally centrifuged, where cell pellets were resuspended in 100 μl of sterile water and plated on selective *-trp*, *-leu* plates for incubation at 30 °C for 2-3 days. Plates lacking tryptophan and leucine are made from a stock solution of 1 L containing 1.2 g of yeast nitrogen base without amino acids and ammonium sulfate, 5 g ammonium sulfate, 10 g succinic acid, 6 g sodium hydroxide, 0.8 g drop-out amino acid mix (-ura, -ade, -trp, -his, -leu), 0.1 g uracil, 0.1 g adenine and 0.05 g his. Co-transformant cultures were grown until OD_{600} = 0.1 in *-trp*, *-leu* liquid medium, where 5 μl of co-transformant cultures were spotted onto *-trp*, *-leu* plates for growth control, and *-trp*, *-leu*, and *-his* plates for interaction analysis, where the *-trp*, *-leu*, and *-his* plates are similar to the *-trp*, *-leu* plates but lack histidine.

3.13 Co-Immunoprecipitation Assays

A 100 μl of cell extract (3 mg/ml) was incubated with 2-3 μl of antibody serum and 24 μg/ml of ethidium bromide for a final volume of 105 μl. The sample was subsequently incubated for 2 hours at 4 °C, while in the meanwhile 50 μl of protein-A sepharose beads were incubated in TBS buffer containing 0.1% Tween-20 (TBS-T). The cell extract containing the antibody serum was added to the equilibrated beads, where another incubation of 2 hours was performed. Three washes with 1 ml of TBS-T were performed, followed by SDS-PAGE (15%) gel running of 15 μl of samples

resuspended in 50 μl of 1 X SDS-PAGE loading buffer and subsequent western blot analysis of the antigen in conjunction with its interactant partner. Western blots with the antibody against RPC34, Rpo13 and Rpo2 were all performed with a 1:1000 dilution of the antibodies.

In order to cross-link the antibodies to the protein-A sepharose beads, 100 μl of bead suspension was vortexed and resuspended in 80 μl of 0.1 M NaPhosphate Buffer (pH 8), with subsequent addition of 30 μl of serum. Beads were rotated for 2 hours at 4 °C, with subsequent centrifuging of the beads and addition of 1 ml of cross-linking buffer (0.2 M triethanolamine, pH = 8.2) containing 25 mM dimethyl pimelimidate (DMP) for 45 min. One ml of blocking buffer was subsequently added (0.1 M ethanolamine, pH 8.2) for 1 h, where a wash with 1 ml of the elution buffer (0.1 M glycine-HCl pH 2.5) was performed to remove all bound antibody that is not cross-linked with DMP. After bead incubation with the cell extract in the presence of ethidium bromide, and washes with TBS-T, the antigen-interactant partner was eluted with 50 μl of 100 mM glycine (pH 2.5) and 0.5 % SDS with sample boiling and subsequent neutralisation with 2 μl of 1 M untitrated Tris.

3.14 'Baby Machine' Cell Synchronisation Procedure

A 'Baby Machine' cell synchronisation procedure was performed as described in [51]. Briefly, a *S. acidocaldarius* culture was grown at $OD_{600} = 0.1$ and added to the 'Baby Machine' apparatus, where cells stick onto the membrane containing the adhesive poly-lysine. Dividing cells falling from the membrane in G1 phase were collected after 3 hours of cell synchronisation procedure in a 50 ml Falcon tube sitting on ice to prevent any growth for a maximum of 15 minutes with a flow rate of 1 ml/min, where a total of 30 ml of synchronised *S. acidocaldarius* culture was grown at 75 °C. Various samples of 2.5 ml volume were collected at various time points for analysis of protein expression levels throughout the *S. acidocaldarius* cell cycle, where 2 ml was centrifuged and resuspended in 25 μl of 1 X SDS-PAGE loading buffer for loading

of 10 μl in SDS-PAGE gels.

3.15 Yeast Protein Cell Extract Preparation

Yeast culture samples of 2 ml were centrifuged, where pellets were resuspended with 300 μl of 20% trichloroacetic acid. Yeast cell lysis was achieved with vortexing in the presence of glass beads, where the protein pellet was resuspended in 200 μl of 1 X high pH SDS-PAGE loading buffer (150 μl of untitrated 1 M Tris base per 1 ml of 1 X buffer). Samples were boiled and centrifuged, with subsequent loading onto SDS-PAGE gels and western blot analysis.

3.16 ChIP-Sequencing

A *S. acidocaldarius* culture was grown in Brock's medium until OD_{600} = 0.4, where formaldehyde was added at a final concentration of 1%. The reaction was quenched with 125 mM glycine, where cell pellets were subsequently collected upon centrifugation at 16,000 g for 20 minutes (4 °C) and washed with 50 ml of 1 X PBS. The cells were resuspended in 4 ml TBS buffer (pH 7.5) containing 0.1% Tween-20 and 0.1% Triton X-100 (TBS-TT). The cell extract was sonicated to generate DNA fragments ranging from approximately 200 bp to 1000 bp (15 minutes total sonication procedure, with 30 seconds ON and 30 seconds OFF), with a subsequent centrifugation step.

Ten μg of cell extract of concentration 205 ng/μl measured by Bradford assay was incubated with 3 μl of antibody-containing serum against Rpo2, Rpo13 and RPC34 raised to a final volume of 100 μl with TBS-TT. Samples were rotated at 4 °C for 2 hours. A volume of 25 μl of protein-A sepharose beads was subsequently added, with rotation at 4 °C for 1 h. Five washes with TBS-TT were performed, with an additional wash with TBS-TT containing 500 mM NaCl and a final wash with TBS-TT containing 0.5% Tween-20 and 0.5% Triton X-100. The protein-DNA complexes were eluted with elution buffer (20 mM Tris, 10 mM EDTA, 0.5% SDS, pH = 7.8),

where protein was removed upon proteinase K digestion. Samples were extracted with a phenol-chloroform step, with subsequent DNA precipitation in ethanol and glycogen. DNA pellets were resuspended in 50 μl TE buffer. The various DNA sequencing libraries were subsequently obtained, where the DNA was processed on an Illumina HiSeq 2000 (Source Bioscience Illumina Sequencing). Reads were mapped to the *S. acidocaldarius* genome using SAMtools and BEDtools to generate .bed files, which were then visualized and quantified using the SeqMonk data analyser (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/).

3.17 RNA Sample Preparation for RNA-sequencing

A 10 ml of *S. acidocaldarius* sample was withdrawn from a culture of OD_{600} = 0.284, where it was subsequently filtered through a 0.45 μM nitrocellulose filter. The filter was subsequently removed, and inserted in a microfuge tube containing 500 μl RNase-free water. The microfuge tube was vortexed to wash the cells off the filter. Lysis buffer (100 μl) was added (600 mM sodium acetate, pH 5.2, 6% SDS) in conjunction with phenol (600 μl) at pH 4.3. Following vortexing and centrifuging, the top aqueous phase was transfered to a new microfuge tube with addition of acid phenol for another round of aqueous phase extraction. 3 M sodium acetate at pH 5.2 (70 μl) in conjunction with isopropanol (750 μl) was added for RNA precipitation. Samples were then treated with 3 μl DNase (50375 U/μl, Invitrogen) at room temperature for 15 minutes, where the integrity of the RNA was checked with an Agilent NanoChip. The RNA sequencing library was subsequently obtained in similarity to the DNA sequencing libraries (Source Bioscience Illumina Sequencing).

3.18 Real-Time Q-PCR

Real-time Q-PCR reactions were performed with the primers described in (Table3.3) for the ChIP-seq experiments performed with the antibodies against Rpo2 and Rpo13, and (Table**??**), (Table**??**) and (Table**??**) for the ChIP-seq experiment

performed with the antibody against RPC34. A SYBR green master mix was used for the Q-PCR reactions (Applied Biosystems).

Gene	Primers
-ve control	**F** CCGTGAACTTGCCTAGAACATTAAAGTATAAGATAATGG **R** CTTTTCCCTTGTTCATAATACCCCTCTTCTTCTTGTTC
Saci_0666 (Thermosome beta subunit)	**F** GTCAGCTACAGCTACAGTTGCTACAACACCTG **R** GTTTATCCTGAGTGCCTCCTTACCATAAGTTC
5S rRNA locus	**F** CCTATATTATGTTGCCCACCCGGTCATAGTG **R** CACGGCCCCTCTGACGTG

Table 3.3: *Primer set used for the Q-PCR reactions on the immunoprecipitated DNA by the antibodies against Rpo2 and Rpo13. The primers were designed against the various genomic loci corresponding to the 5S rRNA locus, the thermosome beta subunit, and a negative region with no ChIP-seq probe value.*

Chapter 4

The Archaeal-Specific Rpo13 Subunit

4.1 Introduction

Downstream DNA enters the RNAP-II through a pair of mobile elements termed jaws [46]. Rbp5 is part of the shelf module, which is the lower jaw [46]. Rbp5 has a bipartite organisation, with an N-terminal domain and a C-terminal globe [182]. The N-terminal module marks the far end of the DNA channel of RNAP-II, and is thought to to mediate Rpb5/DNA contacts found some 15-20 nucleotides ahead of the transcription fork, while the C-terminal part is tightly connected to the Rpb1 subunit [182]. Indeed, Pro86 and Pro118 are exposed in the DNA cleft, and it has been suggested that they insert their cyclic rings near the deoxyribose moieties of downstream DNA [46]. However, they are not (Pro86) or poorly (Pro118) conserved in other eukaryotic phyla, and a double Pro-Thr replacement had no detectable growth defect, suggesting that these DNA contacts, if they occur at all, are of limited biological relevance [182]. On the other hand, a long and highly conserved Rpb5-α1 helix, formed by the first 30 amino acids of Rpb5, was found to be critical *in vivo*. This hydrophilic helix occupies the very end of the lower cleft jaw. Rbp5 is also thought to have a structural role within the RNAP-II. It holds together the

Rpb1-β24/25 and Rpb1-α44/47 folds of RNAPII, which are adjacent to the bridge helix and switch 1 loop respectively [46]. Thus, their binding to Rpb5 connects the bridge helix and switch 1 loop close to the trigger loop and in direct contact with the template DNA strand at the transcription fork [46]. Furthermore, the strong synergy seen between *rpb5* and *rpb9* mutants suggests that Rpb9 closely cooperates with Rpb5 [182]. Indeed, Rpb5 and Rpb9 belong to the upper and lower jaws of the DNA cleft, respectively, and are connected by a large Rpb1 segment which is projected across the DNA cleft [46]. Thus, these Rpb9-Rpb1 and Rpb5-Rpb1 interactions may coordinate the opening/closing of the DNA cleft through Rpb9, and the binding of switch 1 to the transcription fork through Rpb5 [46]. Some transcription factors may assist this conformational change, where TFIIF is thought to bind to Rpb9 [183] as well as to Rpb5 [43]. Rpb5 is also conserved in Archaea in the form of Rpo5. However, the Rpb5 N-terminal module is only found in eukaryotes, while the Rpb5 C-terminal globe is strongly conserved in Archaea and Eukarya [96].

Collaborative work performed by the labs of Dr. Nicola Abrescia and Prof. Stephen Bell has led to the initial structural resolution of the archaeal-specific Rpo13 subunit [96], previously identified by [104]. Rpo13 is only found in the *Crenarchaeota* and *Korarchaeota*, but not in the *Thaumarchaeota* and *Euryarchaeota* [180]. Rpo13 was found to occupy the equivalent position of the eukaryotic-specific Rpb5 jaw domain involved in downstream DNA binding (Figure4.1) [45]. Furthermore, Rpo13 has a helix-turn-helix (HTH) motif with α1 and α2 helices and a third, positively charged C-terminal tail (Figure4.2) [7]. Indeed, Rpo13 has around ten lysines in the last 20 residues, which supports the idea that the third, positively charged C-terminal tail interacts with the negatively charged DNA rather than with the positively charged DNA binding cleft of Rpo1 [96]. Five of these lysines are conserved across Rpo13 of different *Sulfolobales* (Figure4.2). Therefore, Rpo13 might bind to the DNA in an analogous manner to the Rbp5 jaw domain, where the flexible negatively charged N-terminal protein tip as well as the positively charged C-terminal tail could make DNA contacts, embracing it. Furthermore,

Rpo13 could perform some of the equivalent roles of eukaryotic factors, which are absent in Archaea. Moreover, similarly to Rpb5 in RNAP-II, α1 and α2 helices might perform some structural roles, acting as constitutive anchors onto the RNAP. Thus, in an analogous manner to Rpb5, Rpo13 might connect Rpo1's catalytic elements to template DNA in synergy with other RNAP subunits.

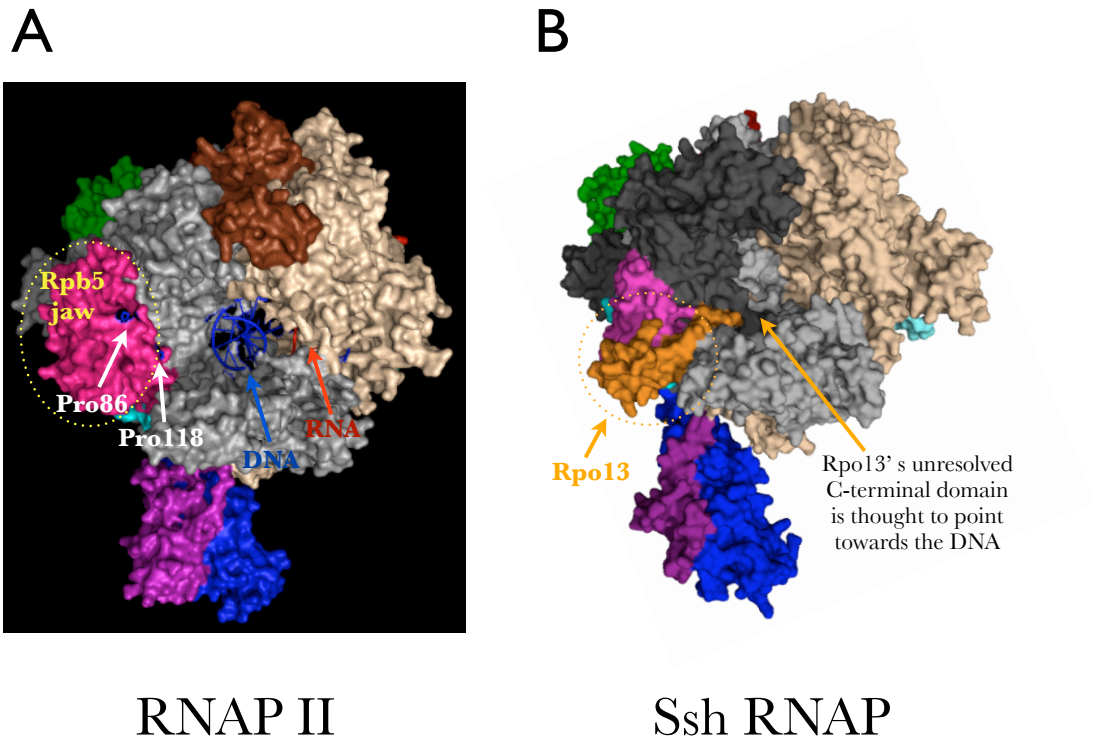


Figure 4.1: *The RNAP-II crystal structure in the presence of DNA in comparison with the archaeal RNAP on its own.* ***(A)*** *The RNAP-II's Rbp5 jaw domain (pink) makes contacts with downstream duplex DNA, where two proline residues (Pro86 and Pro118) have been proposed to be key for its interaction with the DNA.* ***(B)*** *The archaeal RNAP lacks the equivalent of the eukaryotic-specific Rpb5 jaw domain. Instead, a new archaeal-specific Rpo13 subunit has been identified (orange), which occupies the position in the archaeal RNAP equivalent to the Rpb5 jaw domain. Due to this position, Rpo13 is proposed to interact with the DNA in an analogous manner to the Rpb5 jaw domain, where the Rpo13's C-terminal helix, which is poorly resolved but is observed to point towards the DNA, is proposed to be the major determinant of DNA binding. The PDB code for the RNAP-II-DNA crystal structure is 1Y1W, while the PDB code for the archaeal RNAP in the absence of DNA is 2WAQ.*

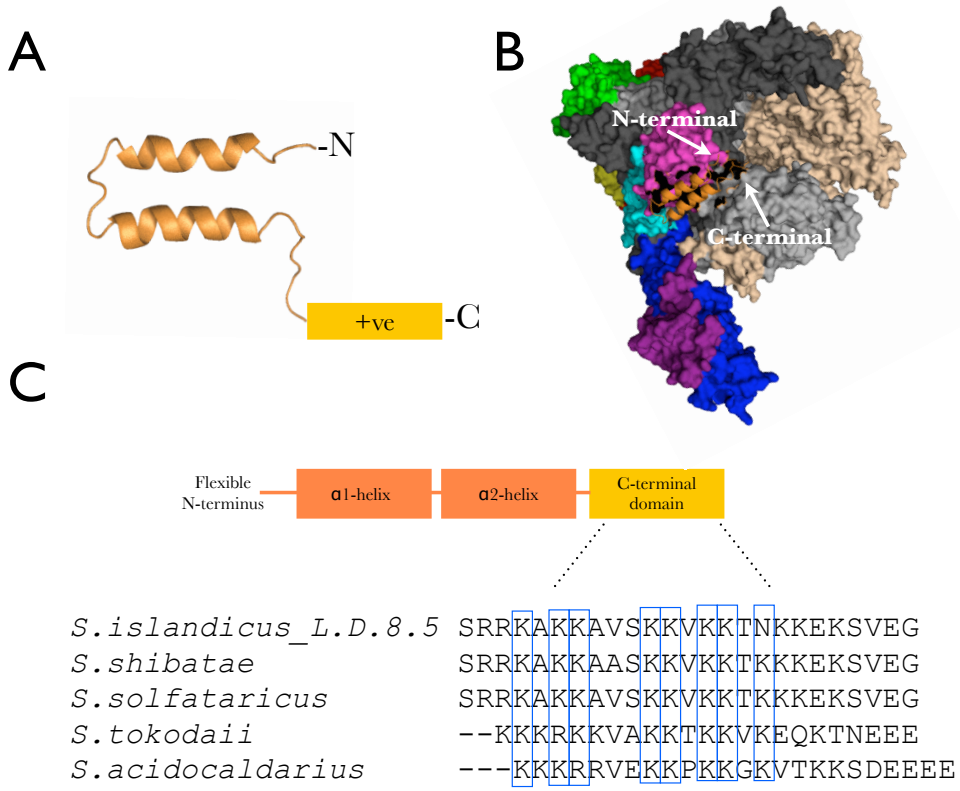


Figure 4.2: *The crystal structure of the positively charged Rpo13 subunit on its own and in the context of the RNAP.* ***(A)*** *Only 45 residues have been resolved, which form an ordered HTH motif fragment with α1 and α2 helices as well as flexible N- and C-termini, where the C-terminal tail is positively charged.* ***(B)*** *Rpo13 in the context of the archaeal RNAP, where the the C-terminal tail is proposed to point towards the DNA in order to interact with it. Rpo13's position in the RNAP analogous to the eukaryotic Rpb5 jaw domain absent in Archaea in conjunction with its positively charged nature supports the idea that it interacts with the DNA.* ***(C)*** *Rpo13 contains about ten lysines in the last 20 residues of the positively charged C-terminal tail, where five of them are conserved across Rpo13 of different Sulfolobales. The PDB code for Rpo13 on its own as well as in the RNAP context is 2WAQ.*

4.2 Results

4.2.1 Rpo13's Cloning and Purification

In order to study the role of Rpo13 from an *in vitro* perspective, plasmids coding for *S. acidocaldarius* and *S. shibatae* Rpo13 were generated as described in Chapter 3 (Materials and Methods). Recombinant *S. acidocaldarius* and *S. shibatae* Rpo13 proteins were subsequently expressed in *E. coli* for purification by S-sepharose cation exchange, where elution was achieved with a shallow NaCl gradient. As can be observed in Figure4.3, pure Rpo13 is obtained, where it is found to migrate at a

higher molecular weight of around 20 kDa than its predicted 12 kDa. The identity of Rpo13 was confirmed by mass spectrometry (MALDI-TOF MS) (Figure4.4). Specifically, a total ion score of 435 for the *S. acidocaldarius* Rpo13 as well as 394 for the *S. shibatae* Rpo13 is obtained, where the total ion score is calculated by weighting ion scores for all individual peptides matched to the protein that is associated with the MS/MS spectrum. The ion score is calculated by the Mascot search engine for each matched peptide from the MS/MS peak lists. This score is based on the probability that ion fragmentation matches are non-random events. If the Ion Score is equal to or greater than the Mascot Significance Level calculated for the search, the protein match is considered to be statistically non-random at the 100% confidence interval.

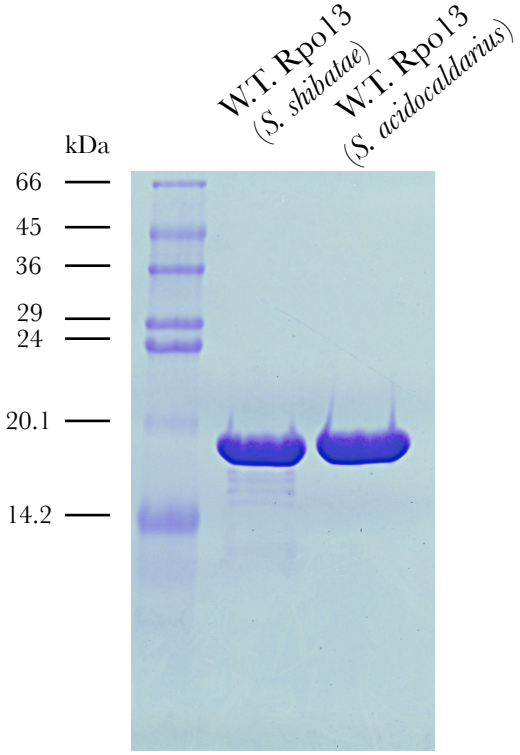
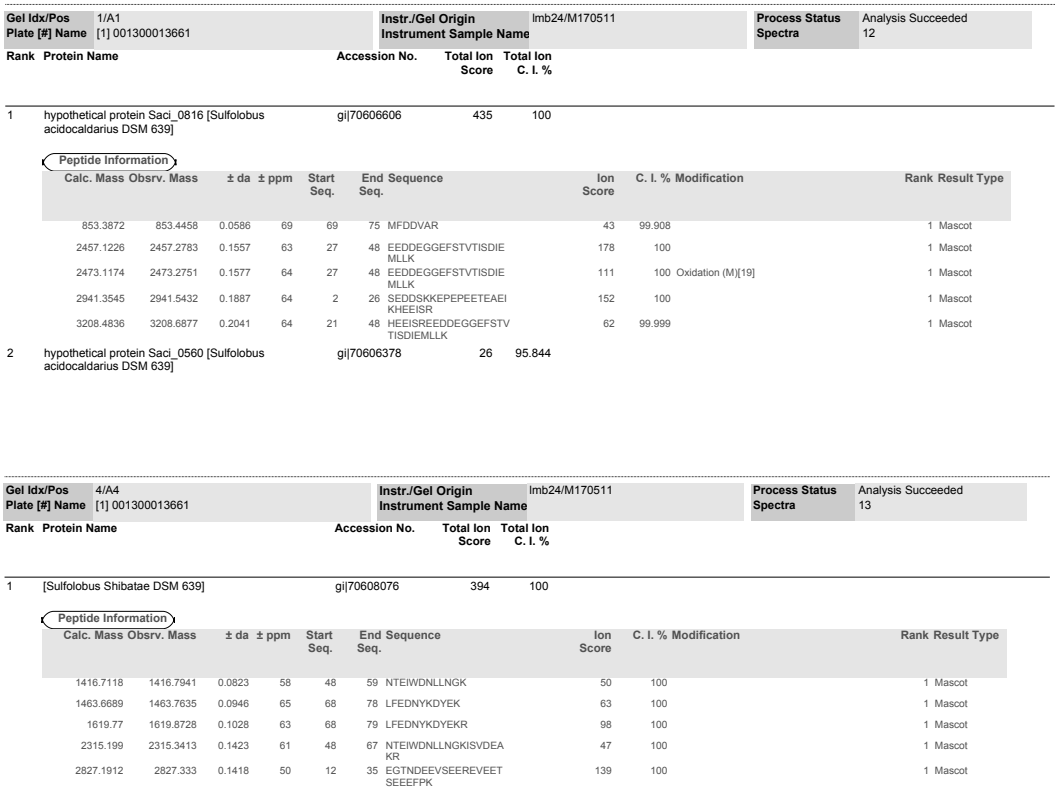


Figure 4.3: *Purified, recombinant Rpo13 from S. acidocaldarius and S. shibatae species, where 20 µg are loaded onto a SDS-PAGE gel subsequently stained with Coomassie Brilliant Blue. Rpo13 is found to migrate at a higher molecular weight of around 20 kDa than its predicted 12 kDa.*



Gel Idx/Pos	1/A1	Instr./Gel Origin	lmb24/M170511	Process Status	Analysis Succeeded
Plate [#] Name	[1] 001300013661	Instrument Sample Name		Spectra	12

Rank	Protein Name	Accession No.	Total Ion Score	Total Ion C. I. %
1	hypothetical protein Saci_0816 [Sulfolobus acidocaldarius DSM 639]	gi|70606606	435	100

Peptide Information

Calc. Mass	Obsrv. Mass	± da	± ppm	Start Seq.	End Seq.	Sequence	Ion Score	C. I. %	Modification	Rank	Result Type
853.3872	853.4458	0.0586	69	69	75	MFDDVAR	43	99.908		1	Mascot
2457.1226	2457.2783	0.1557	63	27	48	EEDDEGGEFSTVTISDIEMLLK	178	100		1	Mascot
2473.1174	2473.2751	0.1577	64	27	48	EEDDEGGEFSTVTISDIEMLLK	111	100	Oxidation (M)[19]	1	Mascot
2941.3545	2941.5432	0.1887	64	2	26	SEDDSKKEPEPEETEAEIKHEEISR	152	100		1	Mascot
3208.4836	3208.6877	0.2041	64	21	48	HEEISREEDDEGGEFSTVTISDIEMLLK	62	99.999		1	Mascot

2	hypothetical protein Saci_0560 [Sulfolobus acidocaldarius DSM 639]	gi|70606378	26	95.844

Gel Idx/Pos	4/A4	Instr./Gel Origin	lmb24/M170511	Process Status	Analysis Succeeded
Plate [#] Name	[1] 001300013661	Instrument Sample Name		Spectra	13

Rank	Protein Name	Accession No.	Total Ion Score	Total Ion C. I. %
1	[Sulfolobus Shibatae DSM 639]	gi|70608076	394	100

Peptide Information

Calc. Mass	Obsrv. Mass	± da	± ppm	Start Seq.	End Seq.	Sequence	Ion Score	C. I. %	Modification	Rank	Result Type
1416.7118	1416.7941	0.0823	58	48	59	NTEIWDNLLNGK	50	100		1	Mascot
1463.6689	1463.7635	0.0946	65	68	78	LFEDNYKDYEK	63	100		1	Mascot
1619.77	1619.8728	0.1028	63	68	79	LFEDNYKDYEKR	98	100		1	Mascot
2315.199	2315.3413	0.1423	61	48	67	NTEIWDNLLNGKISVDEAKR	47	100		1	Mascot
2827.1912	2827.333	0.1418	50	12	35	EGTNDEEVSEEREVEETSEEEFPK	139	100		1	Mascot

Figure 4.4: *The identity of Rpo13 is confirmed upon mass spectrometry analysis, where a a total ion score of 435 for the S. acidocaldarius Rpo13 as well as 394 for the S. shibatae Rpo13 is obtained. All ion scores are either equal to or greater than the Mascot Significance Level calculated for the search, where ion scoring is performed against the Rpo13 protein. Thus, the protein match is considered to be statistically non-random at the 100% confidence interval.*

4.2.2 Antibody Raising against Rpo13

A rabbit antibody was raised against the purified, recombinant *S. acidocaldarius* Rpo13. As shown in figure (Figure4.5), upon incubation of the pre-immune serum control with *S. acidocaldarius* cell extract and pure Rpo13, no bands corresponding to the size of Rpo13 are observed. On the other hand, upon incubation of affinity-purified antibody from the final-bleed serum with *S. acidocaldarius* cell extract and pure Rpo13, equivalent bands corresponding to the pattern of gel migration of Rpo13 of around 20 kDa are observed (Figure4.5), indicating a successful rabbit immune-response against recombinant Rpo13. The affinity-purified antibody raised against *S. acidocaldarius* Rpo13 is found to recognise protein epitopes in *S. shibatae* Rpo13 as

well as a *S. shibatae* Rpo13 mutant lacking the third, positively charged C-terminal tail (Figure4.5), described in the 'Rpo13's mutants purification and characterisation' subsection.

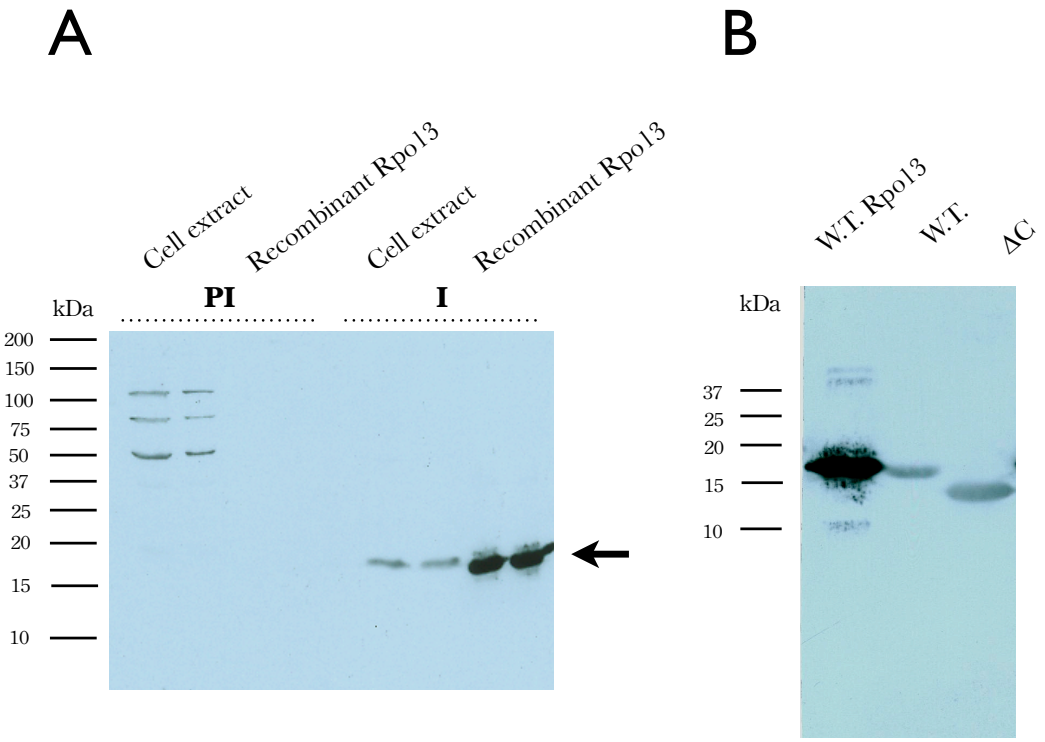


Figure 4.5: *An antibody was raised against recombinant S. acidocaldarius Rpo13.* ***(A)*** *Upon incubation of the pre-immune serum control with S. acidocaldarius cell extract and pure Rpo13, no western blot bands corresponding to the size of Rpo13 are observed, with the exception of background bands. Upon incubation of affinity-purified antibody from the final-bleed serum with S. acidocaldarius cell extract and pure Rpo13, equivalent bands corresponding to the pattern of gel migration of Rpo13 of around 20 kDa are observed, denoted with an arrow.* ***(B)*** *The affinity-purified antibody raised against S. acidocaldarius Rpo13 is found to recognise protein epitopes in S. shibatae Rpo13 as well as a S. shibatae Rpo13 mutant lacking the third, positively charged C-terminal tail (ΔC Rpo13).*

4.2.3 Rpo13's Binding to DNA

In order to test for any DNA binding properties of *S. acidocaldarius* and *S. shibatae* Rpo13, electrophoretic mobility shift assays (EMSAs) were performed with a 40 bp non-specific ssDNA and dsDNA sequence. As can be observed in (Figure4.6), no ssDNA binding activity is observed for the range of protein concentrations tested. On the other hand, several shifted bands are observed when dsDNA is used for the

range of protein concentrations tested (Figure4.6). No shifted bands are observed in the absence of protein. Thus, these shifted bands correspond to protein-DNA complexes. The binding affinity of *S. acidocaldarius* Rpo13 to DNA is calculated to be of an apparent K_d of around 2 μM, where cooperative binding to DNA occurs (Figure4.16).

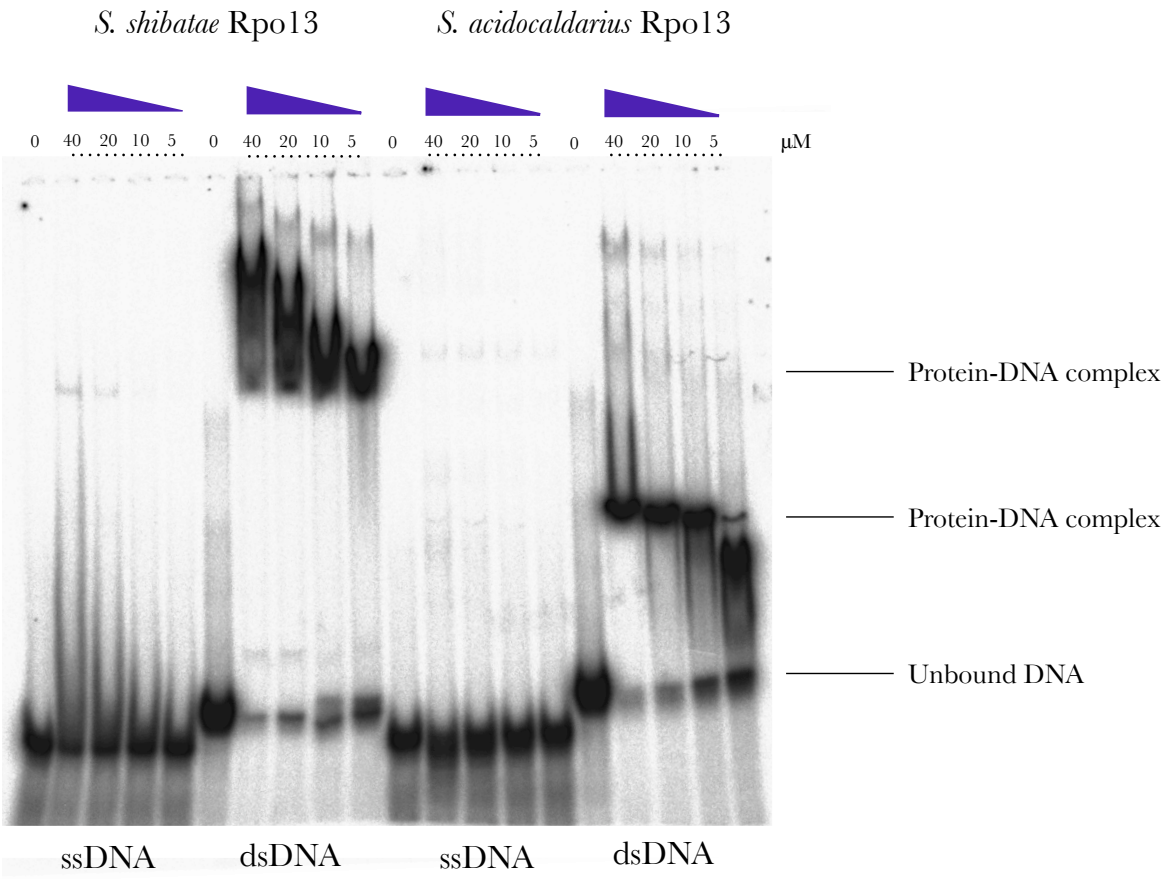


Figure 4.6: *EMSA assays performed with S. acidocaldarius and S. shibatae Rpo13 in the presence of a 40 bp non-specific ssDNA and dsDNA sequence revealed no ssDNA binding activity, while several shifted bands are observed when dsDNA is used for the range of protein concentrations tested. The shifted bands correspond to protein-DNA complexes.*

In order to test for the specificity of Rpo13-DNA complexes as well as Rpo13's identity in the obtained protein-DNA complexes upon EMSA analysis, a supershift approach was performed, where in addition to the incubation of Rpo13 with DNA, the affinity-purified antibody against Rpo13 was also incubated in conjunction with a control antibody against Vps4 [150]. As can be observed in (Figure4.7), upon

incubation of the Rpo13-DNA complexes with the specific antibody against Vps4, no additional bands corresponding to Rpo13-DNA-antibody complexes are observed. On the other hand, upon incubation of the Rpo13-DNA complexes with the affinity-purified antibody against Rpo13, additional bands corresponding to Rpo13-DNA-antibody complexes are observed (Figure4.7). No additional bands are observed upon no addition of an antibody to the Rpo13-DNA complexes.

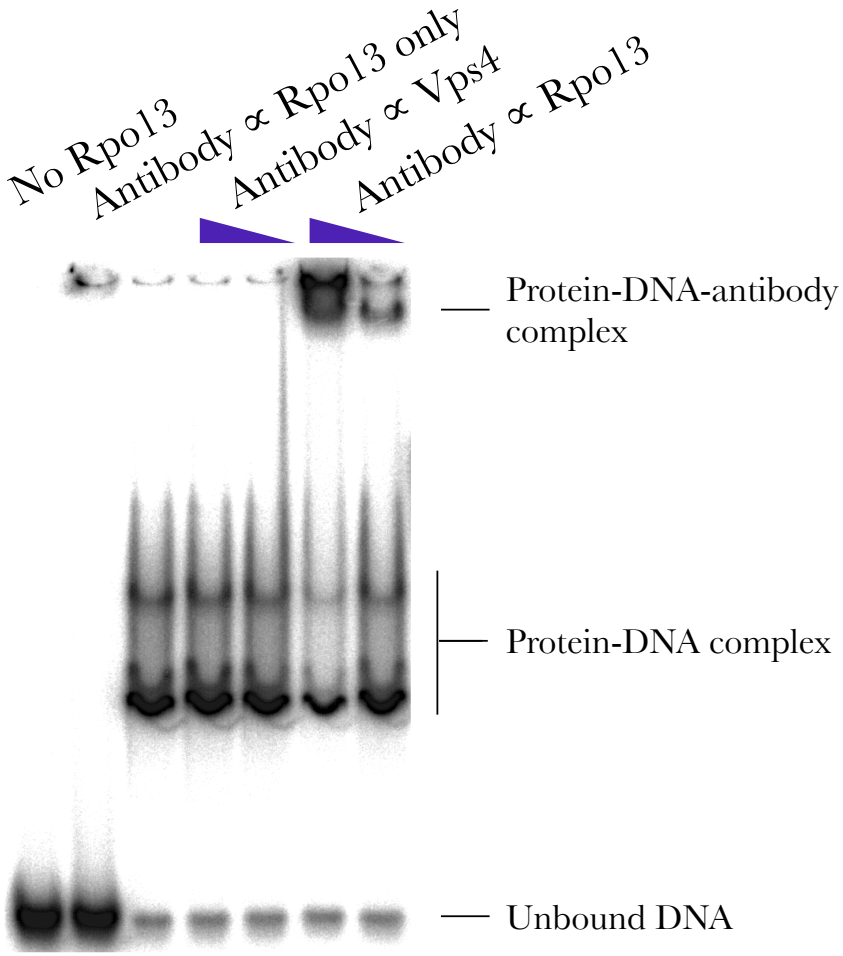


Figure 4.7: *A supershift approach was performed, where S. acidocaldarius Rpo13 was incubated with the 40 bp dsDNA in the presence of either the antibody raised against Vps4 as a negative control [150], or the affinity-purified antibody against Rpo13. Bands corresponding to the protein-DNA-antibody complexes are observed only upon incubation with the affinity-purified antibody against Rpo13.*

In order to complement this supershift analysis to confirm Rpo13's identity in the protein-DNA complexes, a complementary approach was performed. This

approach consisted in performing a standard EMSA procedure, followed by transfer of the protein-DNA complexes to a nylon membrane for DNA retention as well as a nitrocellulose membrane for protein retention, with a subsequent western blot with the affinity-purified antibody against Rpo13. As can be seen in Figure4.8, the *S. acidocaldarius* Rpo13 protein bands are found to migrate at a similar position to the DNA bands at a higher protein concentration, while *S. shibatae* Rpo13 protein bands are found to migrate at a similar position to the DNA bands at a lower protein concentration (Figure4.9). More specifically, the *S. acidocaldarius* and *S. shibatae* Rpo13 protein bands in the presence of DNA were found to migrate at an equivalent position to the Rpo13 protein bands in the absence of DNA in both cases (Figure4.8; (Figure4.9).

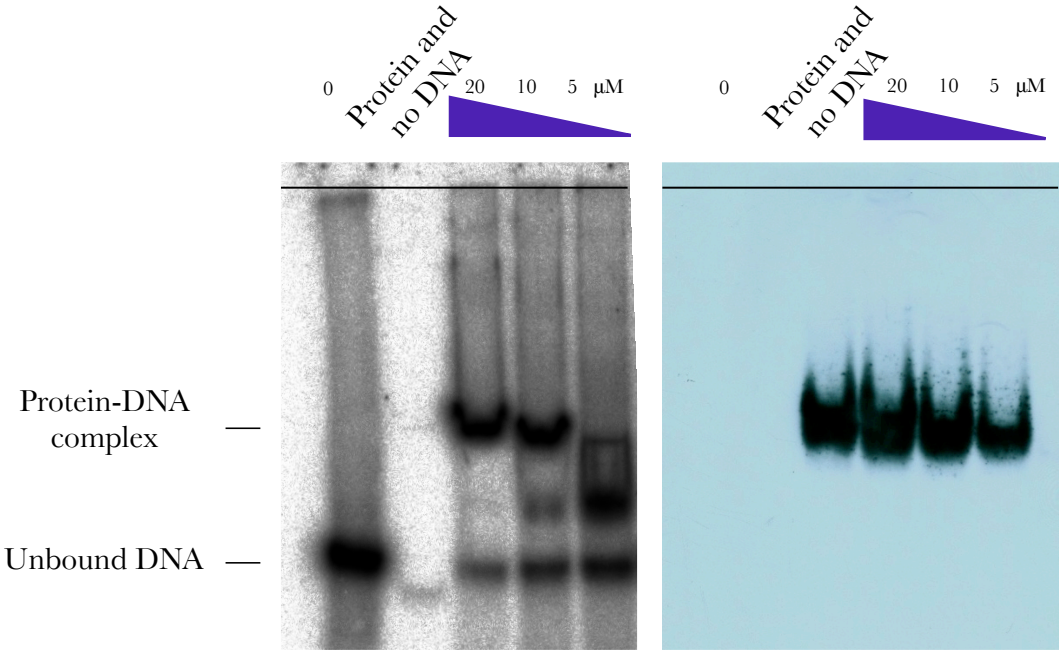


Figure 4.8: *A standard EMSA procedure, followed by transfer of the protein-DNA complexes to a nylon membrane for DNA retention (left) as well as a nitrocellulose membrane for protein retention (right) was performed, where probing against S. acidocaldarius Rpo13 protein in the nitrocellulose membrane was performed with the affinity-purified antibody against Rpo13. The S. acidocaldarius Rpo13 protein bands are found to migrate at a similar position to the DNA bands at a higher protein concentration, where Rpo13 protein bands in the presence of DNA migrate at an equivalent position to the Rpo13 protein band in the absence of DNA.*

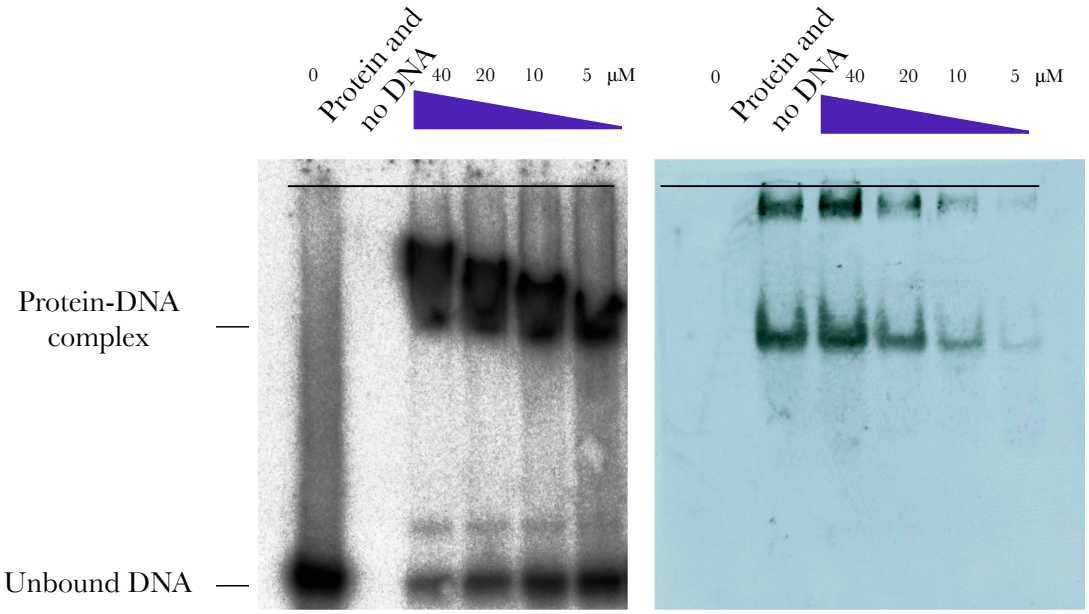


Figure 4.9: *A standard EMSA procedure, followed by transfer of the protein-DNA complexes to a nylon membrane for DNA retention (left) as well as a nitrocellulose membrane for protein retention (right) was performed, where probing against S. shibatae Rpo13 protein in the nitrocellulose membrane was performed with the affinity-purified antibody against Rpo13. The S. shibatae Rpo13 protein bands are found to migrate at a similar position to the DNA bands at a lower protein concentration, where Rpo13 protein bands in the presence of DNA migrate at an equivalent position to the Rpo13 protein band in the absence of DNA.*

4.2.4 The Purification and Characterisation of Rpo13's Mutants

In order to identify the major determinant of Rpo13's binding to the DNA in terms of protein regions, *S. shibatae* Rpo13 mutants lacking either the negatively charged N-terminal tip of Rpo13 (ΔN Rpo13) consisting of the first 33 amino acids, or the the third, positively charged C-terminal tail (ΔC Rpo13) consisting of the last 24 amino acids, or both of these regions (core Rpo13) were purified (Figure4.10). Indeed, from the crystal structure of the archaeal RNAP [96], it is hypothesised that the flexible N- and C-termini mediate DNA binding, with the third, positively charged C-terminal tail being the major determinant of DNA binding. While a standard His-tag affinity purification procedure in the presence of 300 mM NaCl is sufficient for the purification of the ΔC Rpo13 and core Rpo13 mutants, it is found that the salt concentration has to be adjusted to 900 mM NaCl for the purification of the ΔN Rpo13 mutant, since protein aggregation occurs at 300 mM Nacl. Upon

gel filtration analysis with the appropriate salt concentration, it is found that all of the three mutants elute between the void and total volume of the column, in support of the fact that no protein aggregation occurs and that all of the three mutants are stable and folded despite their corresponding missing protein regions (Figure4.11). Furthermore, all of the three mutants are found to elute at an elution volume corresponding approximately to their molecular weights of 9.4 kDa for ΔC Rpo13, 9.7 kDa for ΔN Rpo13 and 6.7 kDa for core Rpo13 (Figure4.11). Thus, pure, stable and properly folded ΔN Rpo13, ΔC Rpo13 and core Rpo13 mutants are obtained.

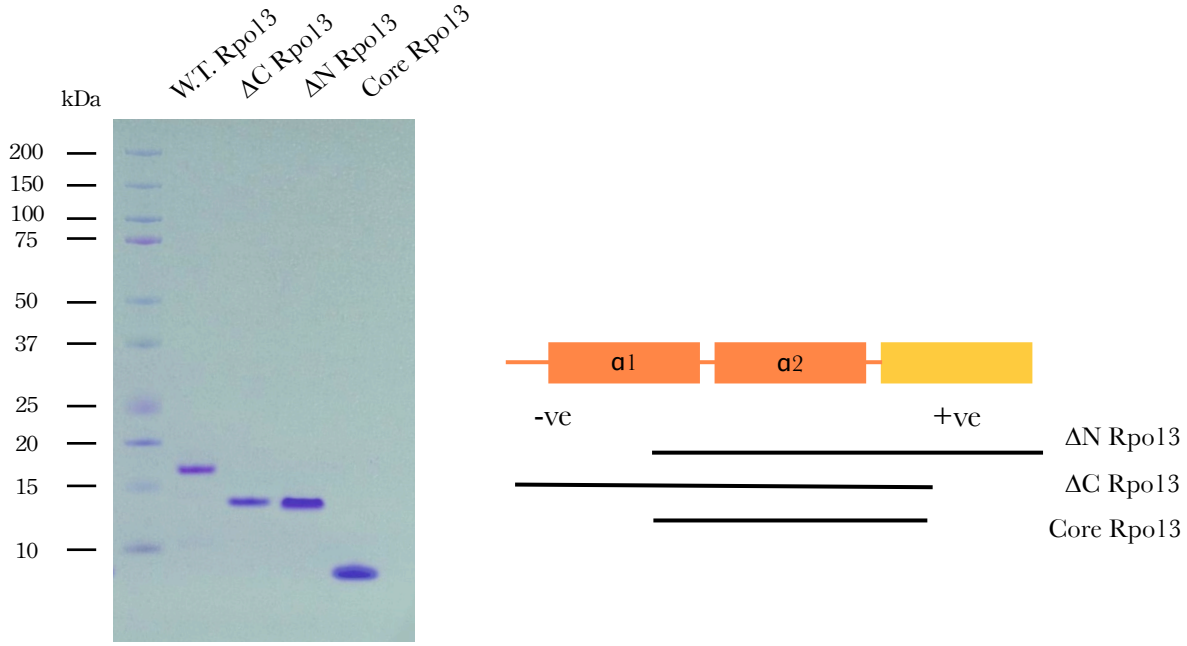


Figure 4.10: *SDS-PAGE gel subsequently stained with Coomassie Brilliant Blue showing purified Rpo13 mutants. These Rpo13 mutants lack either the negatively charged N-terminal tip of Rpo13 (ΔN Rpo13), or the the third, positively charged C-terminal tail (ΔC Rpo13), or both of these regions (core Rpo13). The mutants are found to migrate according to their molecular weights of 9.4 kDa for ΔC Rpo13, 9.7 kDa for ΔN Rpo13 and 6.7 kDa for core Rpo13.*

4.2.5 Analysis of Binding to DNA of Rpo13's Mutants

In order to test for any DNA binding properties of the ΔN, ΔC and core Rpo13 mutants with respect to that of wild-type Rpo13, EMSAs were performed with

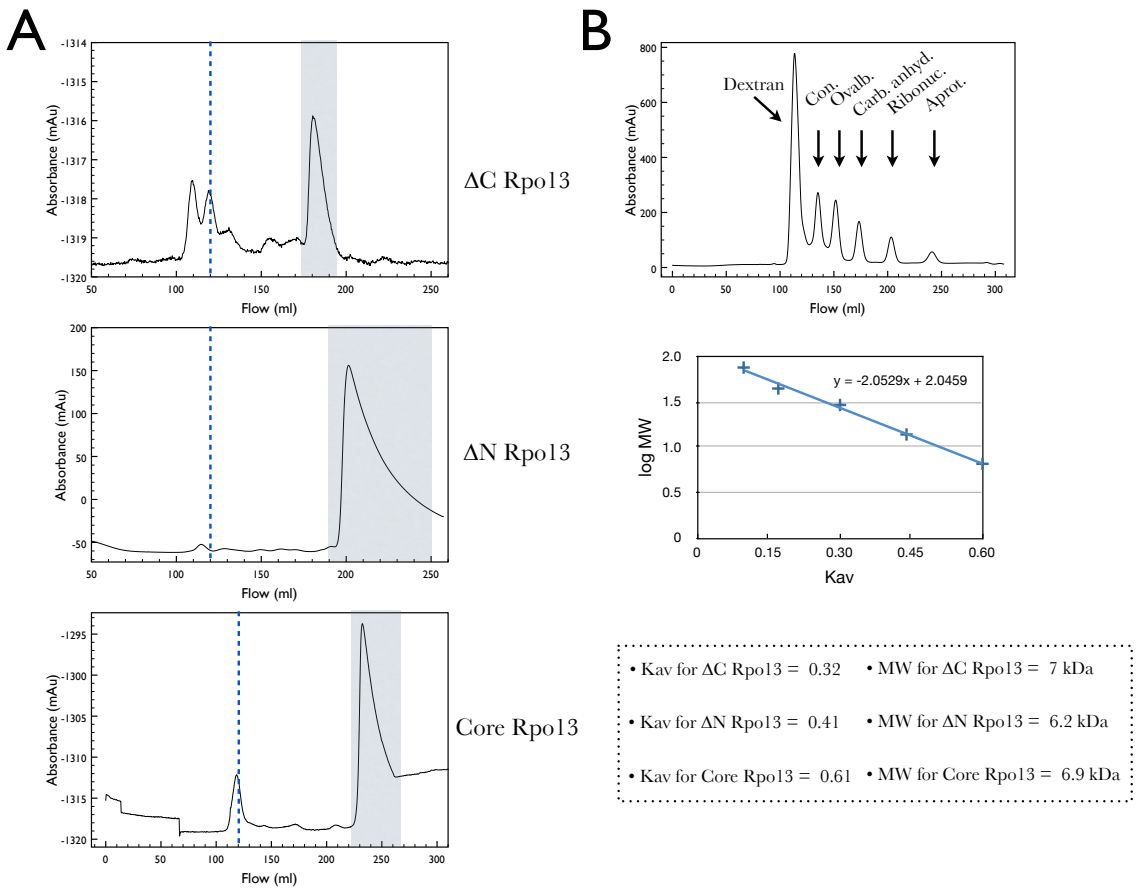


Figure 4.11: *Gel filtration analysis of the Rpo13 mutants.* ***(A)*** *Upon gel filtration analysis of the Rpo13 mutants, it is found that all mutants elute at an elution volume between the void volume of 120 ml (indicated by a blue line) and the total column volume of 350 ml. Specifically, the ΔC Rpo13 mutant is found to elute at an elution volume of 180 ml, while the ΔN Rpo13 mutant is found to elute at an elution volume of 200 ml, and finally the core Rpo13 mutant is found to elute at an elution volume of 240 ml.* ***(B)*** *Protein standards were run onto the gel filtration column: conalbumin (con.) m.w.= 75 kDa, ovalbumin (ovalb.) m.w.= 44 kDa, carbonic anhydrase (carb. anhyd.) m.w.= 29 kDa, ribonuclease A (ribonuc.) m.w.= 13.7 kDa, and aprotinin (aprot.) m.w.= 6.5 kDa. Upon the calculation of the molecular weight of the different Rpo13 mutants according to their elution volumes, it is found that the ΔC Rpo13 mutant has an apparent molecular weight of 7 kDa similar to its true molecular weight of 9.4 kDa, the ΔN Rpo13 mutant has an apparent molecular weight of 6.2 kDa similar to its true molecular weight of 9.7 kDa, and finally the core Rpo13 mutant has an apparent molecular weight of 6.9 kDa similar to its true molecular weight of 6.7 kDa. The differing molecular weight for the ΔN Rpo13 mutant is most likely due to the asymmetrical and irregular gel filtration peak, most likely due to conformational instability of ΔN Rpo13 in solution.*

the same 40 bp non-specific dsDNA sequence as previously used. As can be observed in Figure4.12 and Figure4.13, no dsDNA binding activity is observed for the ΔC Rpo13 and core Rpo13 mutants, as opposed to wild-type Rpo13, which

binds to dsDNA. Also, the residual dsDNA binding activity observed for the ΔC Rpo13 mutant is due to protein contaminants in the ΔC Rpo13 mutant protein preparation, as upon coupling of the EMSA procedure with transfer to nylon and nitrocellulose membranes, no correspondence is found between the DNA and protein bands (Figure4.14). On the other hand, dsDNA binding activity is observed for the ΔN Rpo13 mutant to a level similar to wild-type Rpo13, where incubation reactions were performed with 900 mM NaCl for both proteins (Figure4.15). Furthermore, due to the relative instability of the ΔN Rpo13 mutant at higher protein concentrations, lower protein concentration were used for the dsDNA binding reactions. No shifted bands are observed in the absence of protein. All binding reactions were performed in triplicate (Figure4.16). Thus, while no dsDNA binding activity is observed for the ΔC Rpo13 and core Rpo13 mutants, the ΔN Rpo13 mutant binds to dsDNA to the same degree as wild-type Rpo13.

4.2.6 Rpo13's Binding Mode to DNA

In order to establish Rpo13's binding mode to DNA, that is, whether it binds to the major or minor groove of DNA or to the phosphate backbone, dsDNA binding reactions were performed in the presence of the major groove competitor methyl green as well as the minor groove competitor netropsin. TBP and TFB were used as controls to test whether methyl green and netropsin are functional in their competitor roles when added to dsDNA binding reactions, since TBP and TFB bind to the minor groove of the TATA box and the major groove of the BRE element respectively. As shown in Figure4.17, not only TBP and TFB cooperatively bind to the dsDNA T6 promoter sequence containing a TATA-like A box in the original buffer as previously described [139], [15] as well as in the new buffer used for dsDNA binding of Rpo13 in the presence of the chemicals [155], but also cooperative TBP and TFB binding to the TATA-box containing dsDNA oligo is abolished in the presence of methyl green and netropsin (Figure4.18). This indicates that these chemicals are functional in their competitor roles. On the other hand, no similar

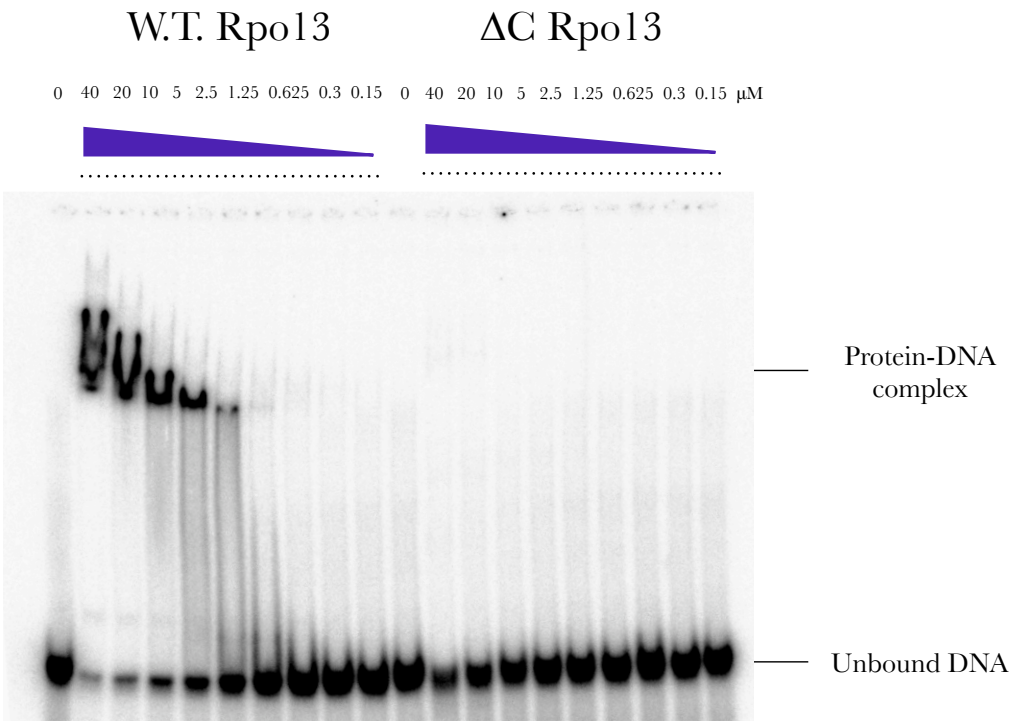


Figure 4.12: *EMSA assays performed with wild-type Rpo13 and ΔC Rpo13 from S. shibatae in the presence of a 40 bp non-specific dsDNA sequence. The assays revealed that, while shifted bands are observed for wild-type Rpo13, corresponding to protein-DNA complexes, for the range of protein concentrations tested, no shifted bands are observed for the ΔC Rpo13 mutant. Thus, the ΔC Rpo13 mutant is not able to bind to the DNA, similarly to wild-type Rpo13.*

effect is obtained upon addition of either methyl green or netropsin in reactions containing Rpo13 incubated with dsDNA in the new buffer that stabilises the chemicals [155], showing that Rpo13 is able to bind dsDNA in the presence of netropsin and methyl green (Figure4.19). Only a minor effect is observed upon netropsin's presence with respect to that of water addition, where slight inhibition of Rpo13's binding occurs, while no significant difference is observed upon methyl green's presence with respect to that of water addition (Figure4.20). Therefore, while methyl green and netropsin abolish cooperative TBP and TFB binding to the DNA, they do not abolish Rpo13's binding to the DNA.

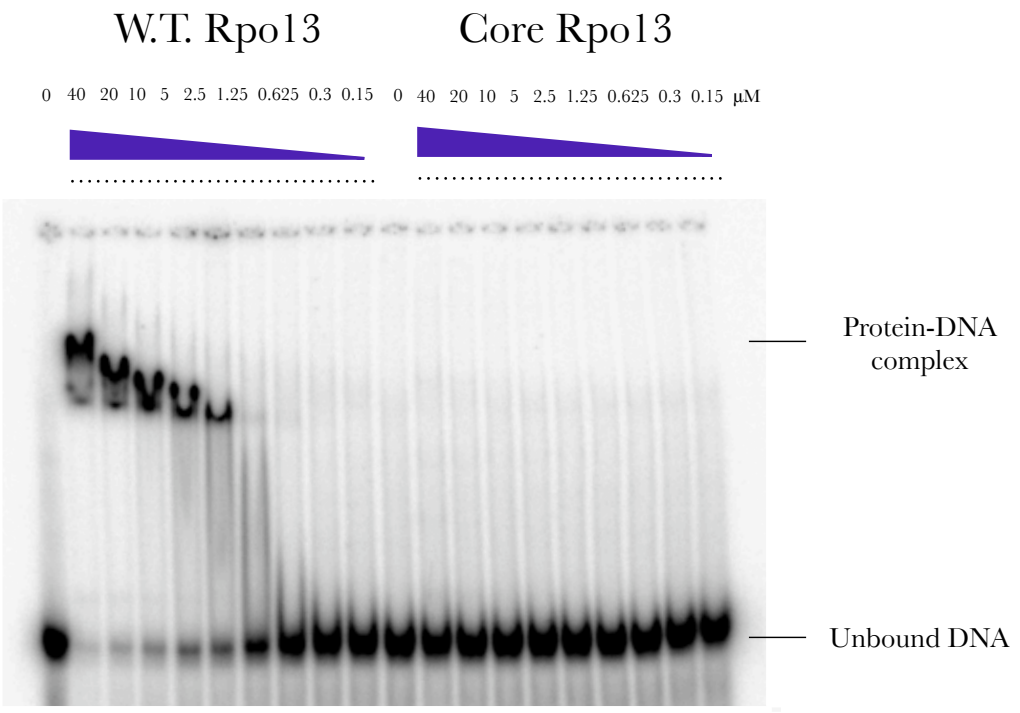


Figure 4.13: *EMSA assays performed with wild-type Rpo13 and core Rpo13 from S. shibatae in the presence of a 40 bp non-specific dsDNA sequence. The assays revealed that, while shifted bands are observed for wild-type Rpo13, corresponding to protein-DNA complexes, for the range of protein concentrations tested, no shifted bands are observed for the core Rpo13 mutant. Thus, the core Rpo13 mutant is not able to bind to the DNA, similarly to wild-type Rpo13.*

4.2.7 Rpo13's Topological Impact on DNA

In order to assess whether Rpo13's binding to dsDNA introduces torsional DNA stress and supercoiling, a plasmid-based supercoiling assay was used [117], [70]. According to this assay, a plasmid containing a single nick on one DNA strand obtained upon incubation with Nb.Bpu10I nicking enzyme is incubated with Rpo13 and then ligated with DNA ligase. If Rpo13 binds to the nicked plasmid DNA in a way to introduce supercoiling, the subsequent ligation of the plasmid would produce various topoisomers. Upon increasing Rpo13's protein concentration, these topoisomers would progressively migrate in an agarose gel lacking ethidium bromide to a position similar to the one of the topoisomer corresponding to supercoiled DNA. As can be observed in Figure4.21, upon incubation of nicked plasmid with no

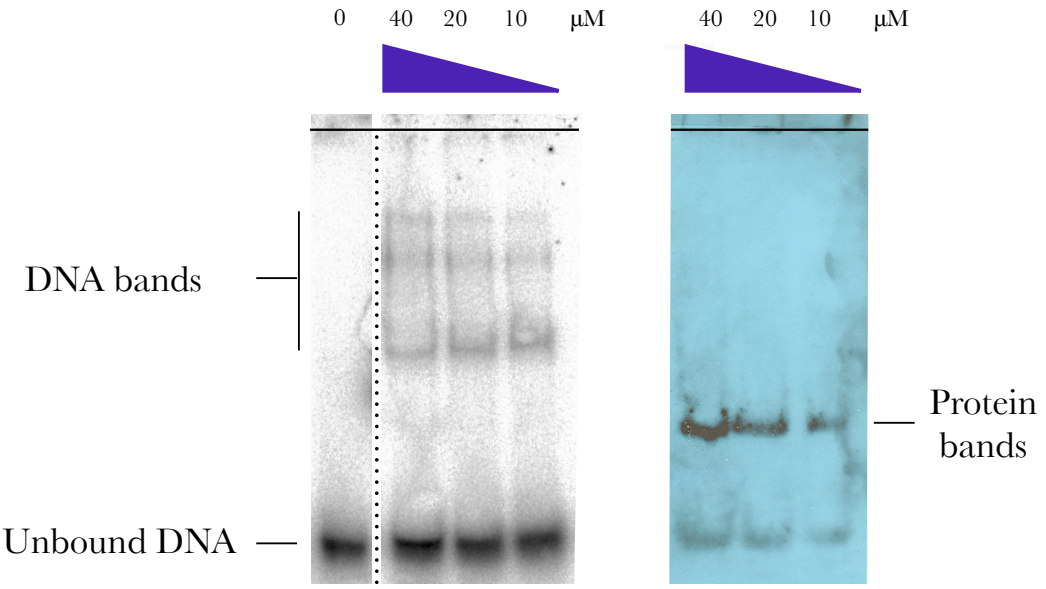


Figure 4.14: *A standard EMSA procedure, followed by transfer of the protein-DNA complexes to a nylon membrane for DNA retention (left) as well as a nitrocellulose membrane for protein retention (right) was performed, where probing against ΔC Rpo13 protein in the nitrocellulose membrane was performed with the affinity-purified antibody against Rpo13. The ΔC Rpo13 protein bands are found to migrate at a different position from the DNA bands, indicating that the protein bands do not co-migrate with the DNA bands.*

Rpo13, various closed circle topoisomers are obtained due to thermal fluctuations. However, upon increasing of Rpo13's protein amount, a change in the distribution of the various topoisomers is observed (Figure4.21). More specifically, the distribution of the various topoisomers is shifted towards the position of migration of the nicked topoisomer upon increasing Rpo13's protein amount, as opposed to the one of the topoisomer corresponding to supercoiled DNA (Figure4.22).

Furthermore, in order to distinguish between the relaxed topoisomer form of the closed circle plasmid versus the nicked topoisomer, which is an open circle plasmid, the various ligation reactions were run on a gel containing ethidium bromide. Indeed, the relaxed topoisomer form of the closed circle plasmid migrates very closely to the nicked topoisomer in an agarose gel containing no ethidium bromide, which may hamper the analysis of topoisomer distribution in the presence of Rpo13. The presence of ethidium bromide would push down the DNA twist, which in turn would push up the DNA writhe and hence the supercoiling of the DNA to either

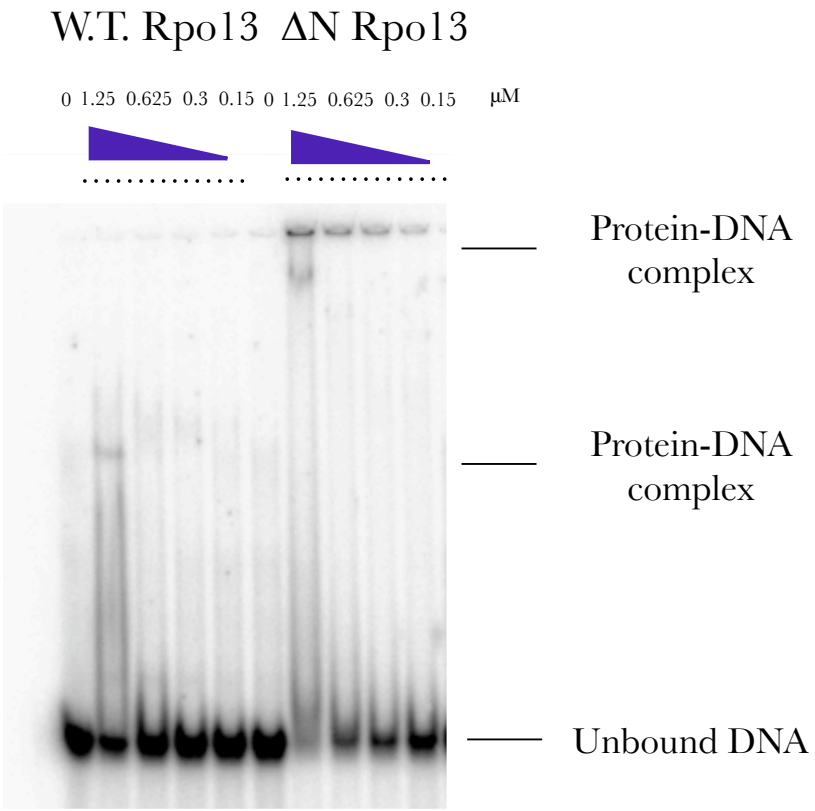


Figure 4.15: *EMSA assays performed with wild-type Rpo13 and ΔN Rpo13 from S. shibatae in the presence of a 40 bp non-specific dsDNA sequence. The assays revealed shifted bands corresponding to protein-DNA complexes for both wild-type Rpo13 and ΔN Rpo13. Lower protein concentrations were used due to ΔN Rpo13's conformational instability in solution and its propensity to aggregation once loaded into the wells of the EMSA gel.*

compact form or positively supercoiled DNA in order to maintain the equivalent linking number. As can be seen in Figure4.23, at lower Rpo13 protein amounts, a DNA band migrating to the position of the one of supercoiled DNA topoisomer is observed in the presence of ethidium bromide. Therefore, the corresponding various topoisomers observed in the absence of ethidium bromide are indeed all closed circle topoisomers (Figure4.23). More importantly, upon increasing Rpo13's amount, the topoisomer distribution is shifted from the supercoiled plasmid position towards the nicked plasmid position in the presence of ethidium bromide (Figure4.23). Thus, in summary, increasing levels of Rpo13 causes a shift in the distribution of topoisomers from closed circle topoisomers towards the nicked topoisomer form.

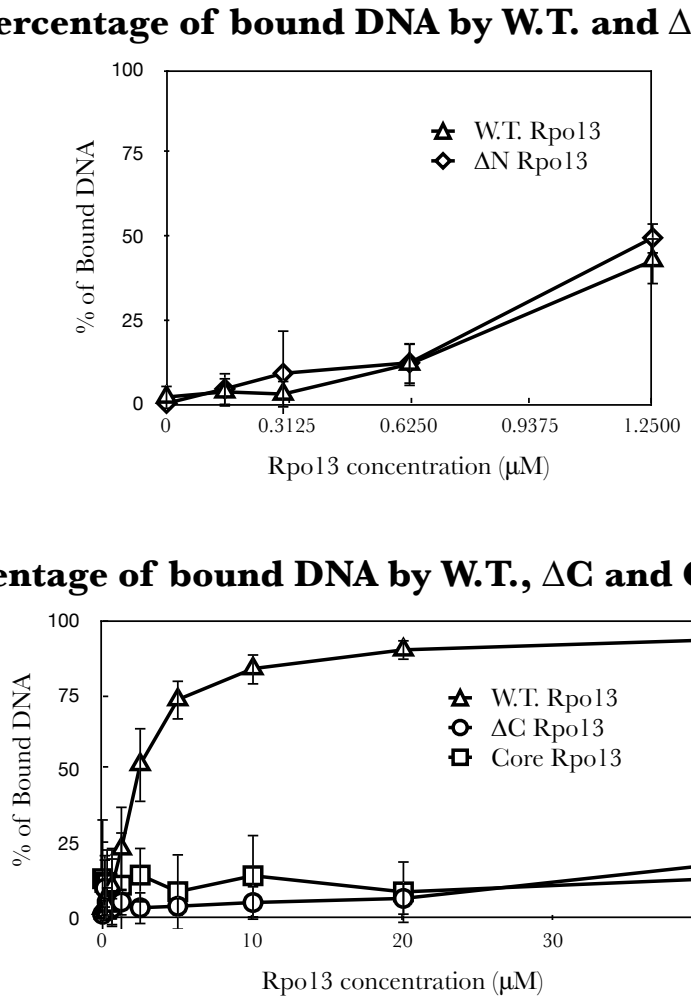


Figure 4.16: *Quantification of % of DNA binding of the three mutants (ΔC Rpo13, ΔN Rpo13 and core Rpo13) with respect to that of wild-type Rpo13, where all EMSA reactions were performed in triplicate. While no dsDNA binding activity is observed for the ΔC Rpo13 and core Rpo13 mutants, the ΔN Rpo13 mutant binds to dsDNA to the same degree as wild-type Rpo13. Lower protein concentrations were used for the ΔN Rpo13 mutant due to its conformational instability, where the concentration of wild-type Rpo13 was also lowered to the one of ΔN Rpo13. The K_d for wild-type Rpo13's binding to DNA is calculated to be of around 2 μM.*

4.2.8 Rpo13's *in vivo* Association with the Archaeal RNAP within the *S. acidocaldarius* Cell Extract

In order to analyse the *in vivo* association of Rpo13 with the archaeal RNAP, a gel filtration procedure was undertaken with a Superdex-200 column, where the elution volumes of purified RNAP from the *S. acidocaldarius* cell extract and recombinant

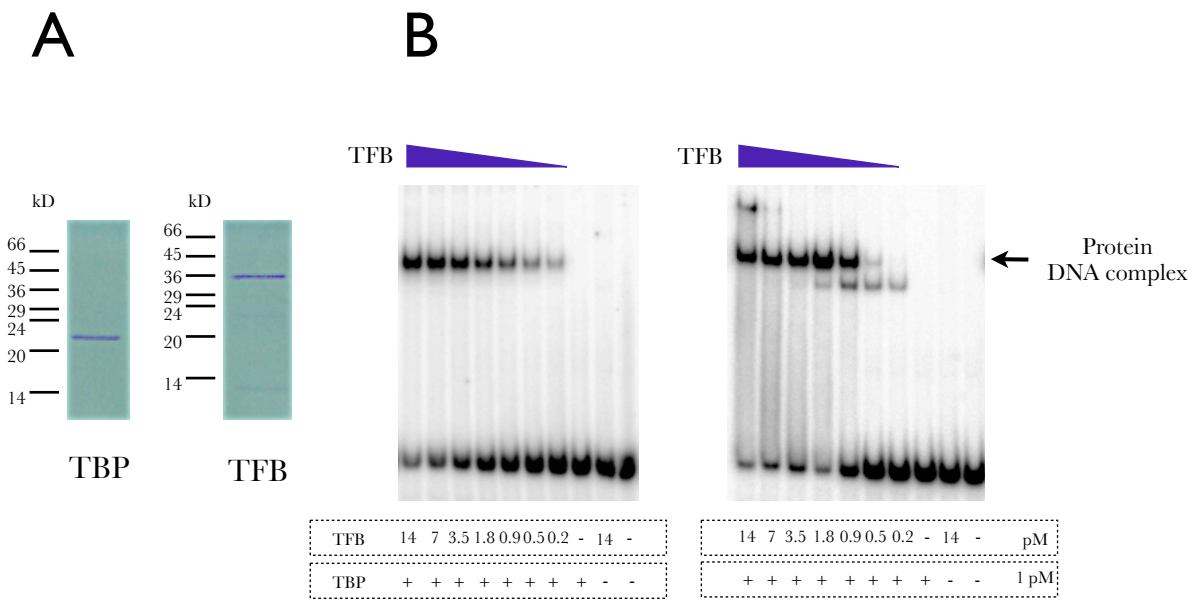


Figure 4.17: *Cooperative binding of TBP and TFB to DNA.* ***(A)*** *Purified TBP and TFB, which run in the SDS-PAGE gel according to their molecular weights of 19 kDa and 35 kDa respectively.* ***(B)*** *(Left) Cooperative binding of TBP and TFB to the T6 promoter sequence containing the TATA-like A box in the original buffer, and (right) cooperative binding of TBP and TFB to the T6 promoter sequence containing the TATA-like A box in the new buffer that solubilises methyl green and netropsin. No DNA binding is observed when only either TBP or TFB are added individually.*

Rpo13 were compared to those obtained from running *S. acidocaldarius* cell extract. More specifically, as Figure4.24 shows, the elution volume of recombinant Rpo13 was very distinct from that of purified RNAP, monitored with an antibody against Rpo2 (120 kDa) and Rpo13 (20 kDa). Both proteins eluted at a volume corresponding to their molecular weight of 44 kDa for Rpo13 which forms dimers in solution [179] and 450 kDa for the RNAP. Furthermore, the *S. acidocaldarius* cell extract was run, where ethidium bromide was added in order to disrupt any protein interactions mediated by DNA. Subsequently, western blot analysis was performed to visualise Rpo2 and Rpo13 across a range of elution volumes containing RNAP and Rpo13. As can be seen in Figure 4.24, not only is Rpo13 found to co-elute with Rpo2 across an elution volume corresponding to RNAP, but it is also found to elute at a different elution volume corresponding to recombinant Rpo13. Importantly, all Rpo13 co-elutes with Rpo2 when purified RNAP is run in the gel filtration

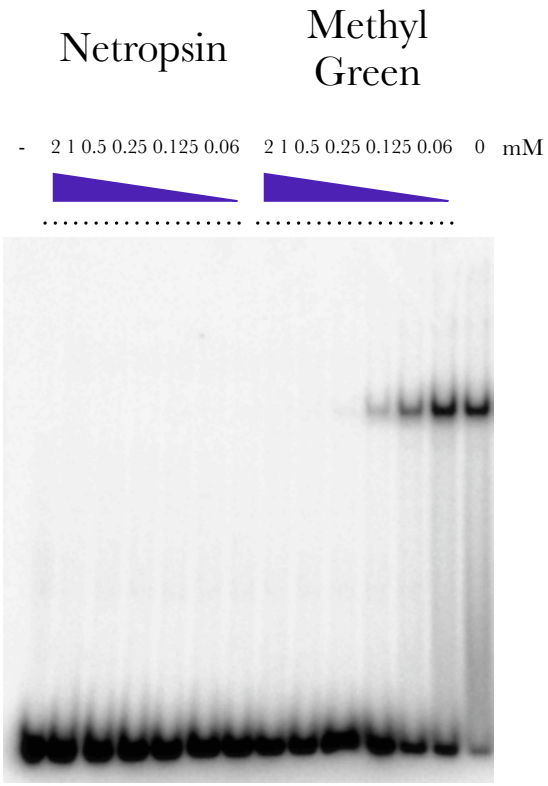


Figure 4.18: *Cooperative binding of TBP and TFB to the T6 promoter sequence containing the TATA-like A box is abolished when netropsin is added, which competes against TBP for minor groove binding, and when methyl green is added, which competes against TFB for major groove binding. Cooperative binding of TBP and TFB is observed when no competitor is added (0 mM). The constant amount of proteins added in all lanes is of 7 pM for TFB and 1 pM for TBP.*

column, which excludes any eventual Rpo13's dissociation event from the RNAP in the *S. acidocaldarius* cell extract during the gel filtration procedure. Thus, in summary, Rpo13 within the *S. acidocaldarius* cell extract is mainly associated with the endogenous RNAP since it co-elutes with its Rpo2 subunit, while there is a small Rpo13 pool which is not associated with the RNAP and elutes at an elution volume corresponding to recombinant Rpo13.

4.3 Discussion

The archaeal-specific Rpo13 subunit occupies a position in the archaeal RNAP which, in RNAP-II, is filled by the eukaryotic-specic Rpb5 jaw domain. It has

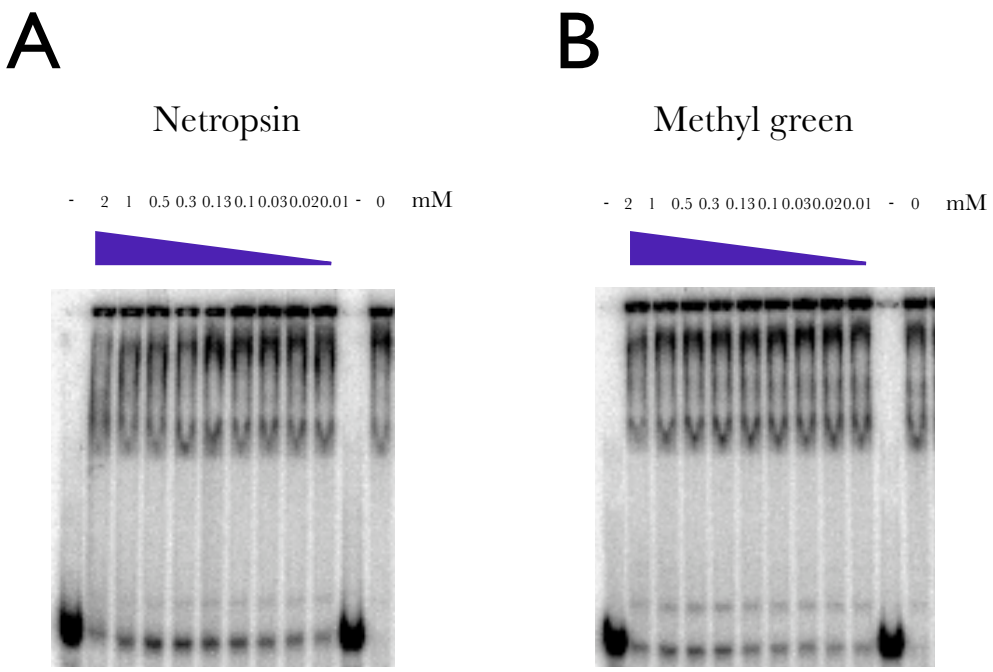


Figure 4.19: *Competitive binding of Rpo13 in the presence of either netropsin or methyl green.* ***(A)*** *Competitive binding of Rpo13 in the presence of netropsin, where only slight inhibition of Rpo13's binding to the DNA is observed with respect to that of the binding reaction where no netropsin is added (0 mM).* ***(B)*** *Competitive binding of Rpo13 in the presence of methyl green, where no inhibition of Rpo13's binding to the DNA is observed with respect to that of the binding reaction where no methyl green is added (0 mM). A constant Rpo13 concentration of 5 μM is added for each lane.*

therefore been proposed to interact with DNA in an analogous manner to Rb5. Specifically, while Pro86 and Pro118 residues and a highly conserved Rpb5-α1 helix, which occupies the very end of the lower cleft jaw, are thought to be important for Rbp5's interaction with DNA ([46], [182]), the flexible N- and C-termini of Rpo13 have been proposed to mediate DNA binding, with the third, positively charged C-terminal tail being the major determinant of Rpo13's binding to DNA [96].

In order to study the role of Rpo13 from an *in vitro* point of view, recombinant *S. acidocaldarius* and *S. shibatae* Rpo13 proteins were purified with a S-sepharose cation exchange step, since Rpo13 has a positively charged C-terminal tail. Rpo13 was found to migrate in an SDS-PAGE system at a higher molecular weight than its predicted 12 kDa. This is most likely due to the retention of migration in the

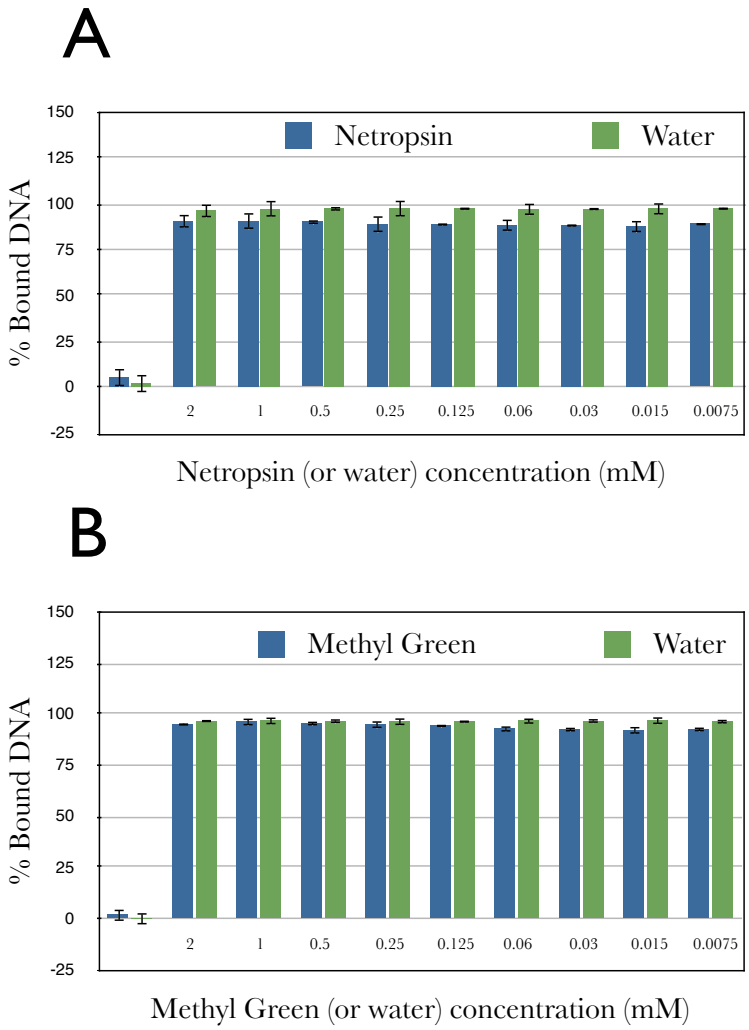


Figure 4.20: *Quantification of Rpo13's binding to the DNA in the presence of either netropsin (**(A)**) or methyl green (**(B)**). While only slight inhibition of Rpo13's binding to the DNA is observed in the presence of netropsin with respect to that of the binding reaction where water is added, no inhibition of Rpo13's binding to the DNA is observed in the presence of methyl green with respect to that of the binding reaction where water is added. All EMSA reactions were performed in triplicate.*

SDS-PAGE system by the third, positively charged C-terminal tail. Nonetheless, Rpo13's identity was confirmed by mass spectrometry (MALDI-TOF MS).

Upon incubation of *S. acidocaldarius* and *S. shibatae* Rpo13 with a 40 bp non-specific ssDNA and dsDNA sequence, protein-DNA complexes were observed only in the presence of dsDNA. Thus, Rpo13 is specific for dsDNA only. The different migration patterns are most likely due to the different pIs of *S. acidocaldarius* and

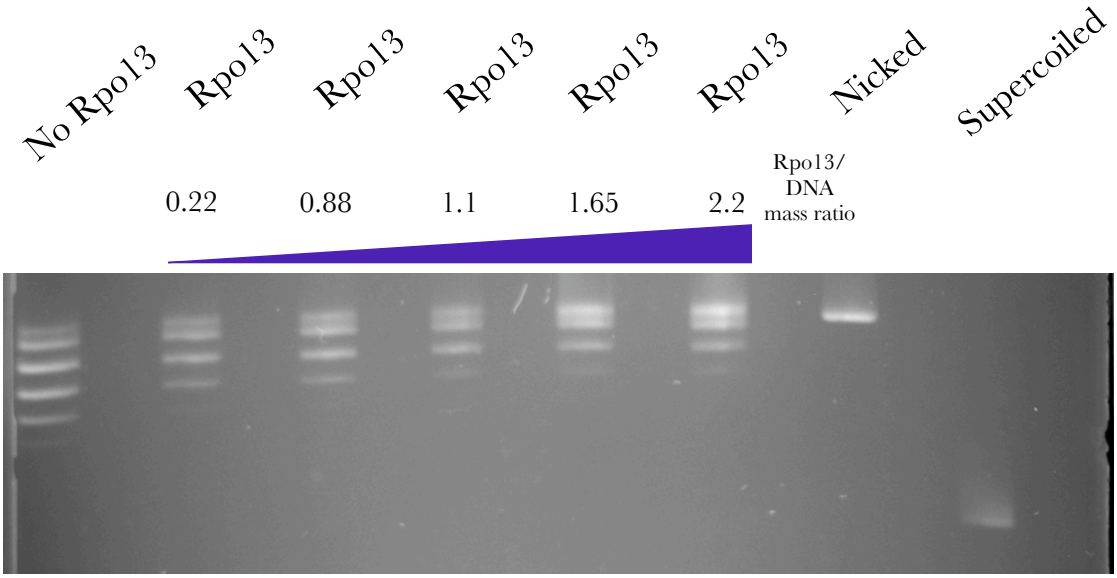


Figure 4.21: *A plasmid-based supercoiling assay, where Rpo13 is incubated at different protein/DNA mass ratios with a constant amount of 0.5 µg pBR322 plasmid. Various plasmid topoisomers are observed in the absence of Rpo13, due to thermal fluctuations. As Rpo13's protein amount is increased, the distribution of the various topoisomers is shifted to the one corresponding to the nicked topoisomer, as opposed to the one corresponding to supercoiled DNA. Agarose gel staining was performed with SYBR Green I staining.*

S. shibatae Rpo13, with *S. acidocaldarius* Rpo13 having a pI value of 4.4 and thus a faster migration rate, while *S. shibatae* Rpo13 has a pI value of 6.1. Also, the binding affinity of *S. acidocaldarius* Rpo13 to DNA was calculated to be of an apparent K_d of around 2 μM, where the weak DNA binding affinity is compatible with Rpo13 not compromising transcriptional processivity but rather modulating the interaction with DNA with no sequence specificity.

In addition, at higher *S. shibatae* and *S. acidocaldarius* Rpo13 protein concentrations, complexes were found to migrate at a higher position than the complexes formed at lower Rpo13 concentrations (Figure4.6). This indicates that *S. shibatae* Rpo13 displays multiple protomer DNA binding as well as *S. acidocaldarius* Rpo13 having a higher dissociation rate at lower protein concentrations. Indeed, upon EMSA analysis coupled with a western blot, it can be observed that *S. shibatae* Rpo13 is capable of forming a higher protein complex, perhaps representing a dimer. Collaborative work performed by Dr. Nicola Abrescia showed that upon multiangle

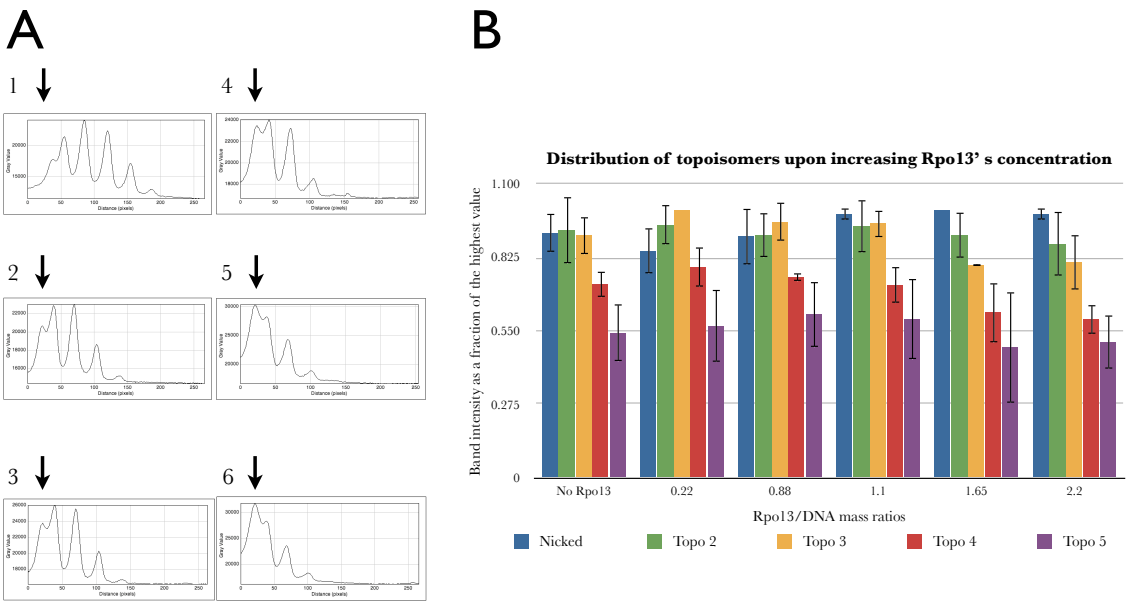


Figure 4.22: *Quantification of the distribution of the various topoisomers upon increasing Rpo13's amount.* ***(A)****) Quantification of the band intensity of the various topoisomers upon no incubation with Rpo13 (1) and increasing Rpo13's amount (2-6). Agarose gel staining was performed with SYBR Green I staining.* ***(B)****) Quantification of the band intensity for the various topoisomers obtained with no incubation with Rpo13 and increasing protein/DNA mass ratios. All values are expressed as a fraction of the highest value. Supercoiling assays were performed in triplicate.*

laser light scattering analysis of *S. shibatae* Rpo13, *S. shibatae* Rpo13 forms dimers in solution [179]. Furthermore, both supershift analysis performed with the antibody raised against Rpo13 as well as the EMSA analysis coupled with a western blot showed that DNA binding is mediated by Rpo13. These antibody-Rpo13-DNA complexes are observed as well as co-migration of Rpo13 with DNA. Interestingly, for both *S. shibatae* and *S. acidocaldarius* Rpo13, western blot protein bands in the presence of DNA were found to migrate at the same position as those of protein bands without DNA, suggesting that the strong positive and negative charged regions of Rpo13 dictate the overall pattern of migration in the presence of DNA. Also, multimeric *S. shibatae* Rpo13 binding might occlude the epitope for antibody binding, explaining why at higher protein concentrations, DNA bands migrate at a higher position than western blot protein bands. Alternatively, dissociation of protein-DNA complexes might occur during gel migration.

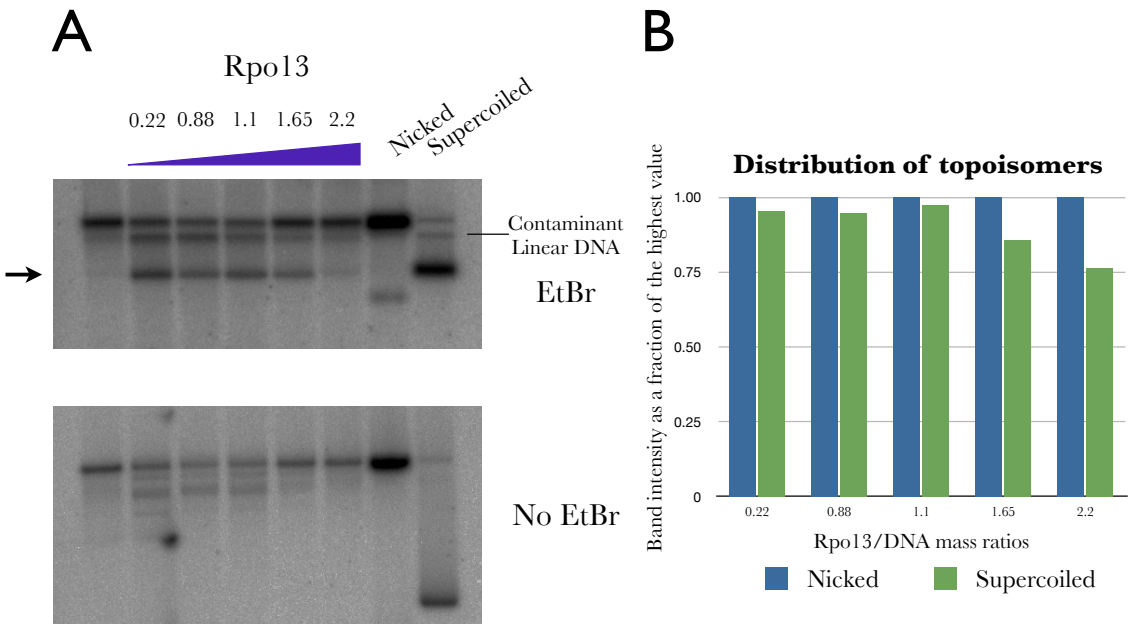


Figure 4.23: *Supercoiling assays performed in the presence of ethidium bromide, with a subsequent Southern blot procedure.* ***(A)****) The same various supercoiling reactions were divided into two, and run either in the presence of ethidium bromide (top), or without ethidium bromide (bottom). At lower Rpo13's protein amounts, a DNA band migrating to the position of the one of supercoiled DNA topoisomer is observed in the presence of ethidium bromide (indicated by an arrow). Thus, the corresponding various topoisomers observed in the absence of ethidium bromide are all closed circle topoisomers.* ***(B)****) Upon increasing Rpo13's amount, the topoisomer distribution is shifted from the one of supercoiled plasmid towards the one of nicked plasmid.*

It has been proposed that Rpo13 binds to DNA, by embracing it. This is maybe due to Rpo13's helix-turn-helix (HTH) motif with $\alpha 1$ and $\alpha 2$ helices and a third, positively charged C-terminal tail as well as a flexible negatively charged N-terminal tip. Specifically, the third, positively charged C-terminal tail is most likely the major determinant of DNA binding, while the negatively charged N-terminal tip might make only minor contacts. In order to test the roles of both the N- and C-termini in DNA binding, mutants lacking either the negatively charged N-terminal tip (ΔN Rpo13), or the positively charged C-terminal tail (ΔC Rpo13), or both of these regions (core Rpo13) were purified. While both of the ΔC Rpo13 and core Rpo13 mutants could be purified with a standard His-tag affinity purification procedure in the presence of 300 mM NaCl, it was found that the salt concentration has to be adjusted to 900 mM NaCl for the purification of the ΔN Rpo13 mutant, since protein

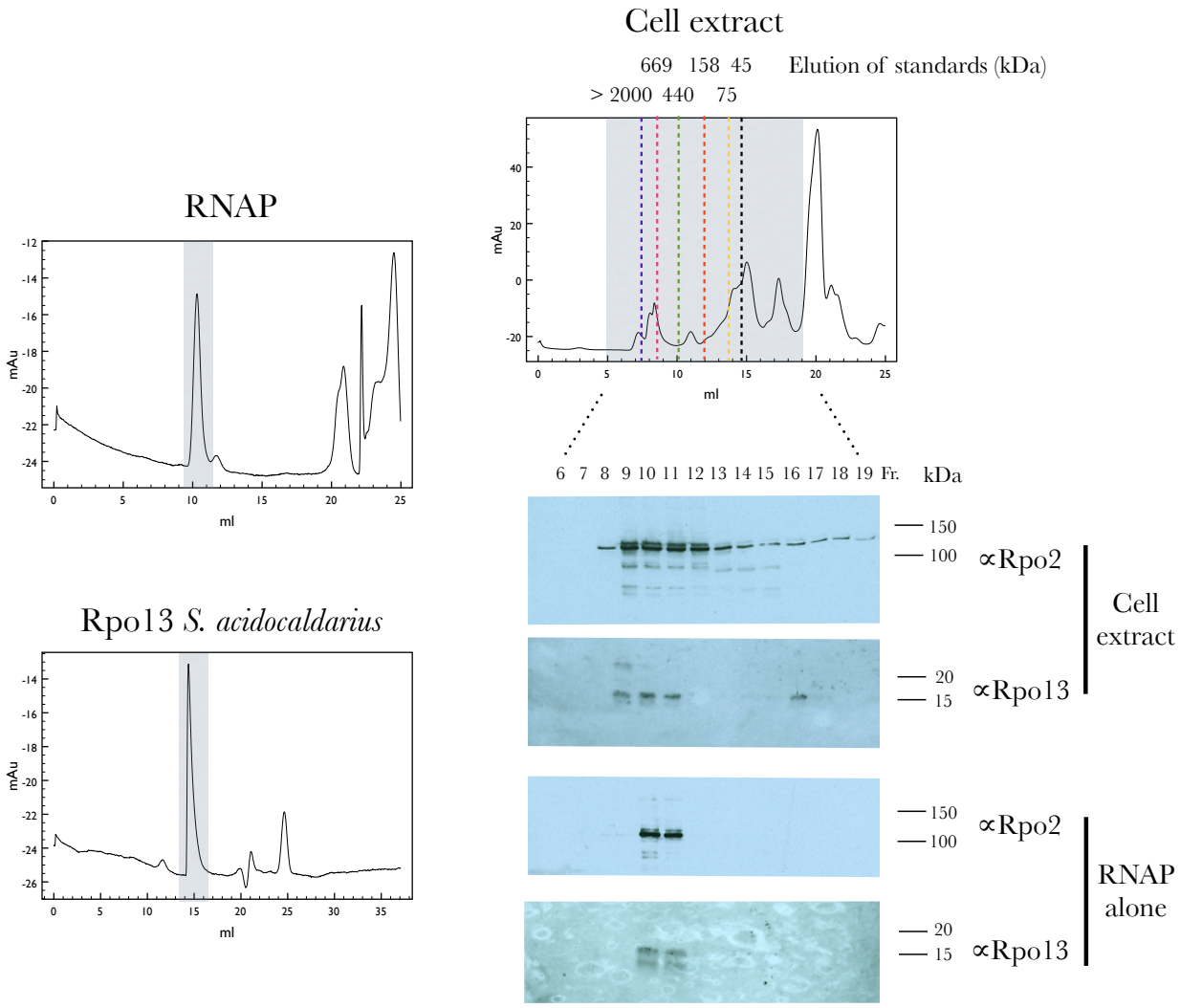


Figure 4.24: *Purified RNAP from the S. acidocaldarius cell extract and Rpo13 elute respectively at 10 ml and 15 ml, which corresponds to their molecular weights of 450 kDa and 44 kDa, which accounts for Rpo13's ability to form dimers in solution. Upon analysis of the RNAP from the S. acidocaldarius cell extract , it is found that Rpo13 co-elutes with Rpo2 across an elution volume corresponding to the one of the RNAP. Furthermore, Rpo13 elutes also at a different volume of around 15-16 ml, which corresponds to the elution volume of recombinant Rpo13. Finally, all Rpo13 co-elutes with Rpo2 when the RNAP is run, excluding the idea that Rpo13 dissociates from the bulk RNAP pool in the cell extract during the gel filtration procedure.*

aggregation occurs at 300 mM NaCl. Indeed, at low salt concentration, the absence of the negatively charged N-terminal tip caused a strong electrostatic repulsion of the positively charged C-terminal tail in solution, thus resulting in protein aggregation. On the other hand, upon salt addition, a salting-in effect occurred, where ions from the salt associate with the ΔN Rpo13 protein's surface, in particular the positively

charged C-terminal tail. This shields these areas from electrostatic, intramolecular repulsion as well as water molecules, and as a consequence, less water molecules are required to interact with the ΔN Rpo13 protein's surface. The concentration of free water was therefore increased, increasing the solubility of the ΔN Rpo13 protein in solution. Upon DNA binding analysis, it was found that neither the ΔC Rpo13 nor core Rpo13 mutants bind to DNA, implying that the third, positively charged C-terminal tail is the major determinant of DNA binding. On the other hand, the ΔN Rpo13 mutant was found to bind to DNA to the same degree as wild-type Rpo13, with a slower migration than the one of wild-type Rpo13 due to a higher pI value of 9.7. Therefore, the flexible N-terminal region of Rpo13 is not required for Rpo13's binding to DNA. In summary, as shown in Figure4.25, the positively charged C-terminal tail of Rpo13 is the major determinant of DNA binding, while the N-terminal tip is not required but may nonetheless help mediating some DNA contacts. The EMSA biochemical data showing Rpo13's binding to DNA is supported by the first crystal structure of the open-complex of the archaeal *S. shibatae* RNAP in the presence of bubble DNA oligo obtained by collaborative work performed by Dr. Nicola Abrescia (Figure4.25) ([180]). This is an improvement of the previous crystal structure ([96]) in that the DNA is present in the RNAP and the structure of Rpo13 is resolved at a higher resolution. Indeed, this RNAP crystal structure shows that the N-terminal Rpo13 tail, with the first visible Cα N-terminal Phe33 residue as well as the Glu32 residue, pointing towards the DNA, and the poorly resolved, flexible C-terminal tail also facing in the direction of contact with the DNA. In addition, while α1 helix makes amino acid contributions with Rpo5 and the clamp-head domain of Rpo1N, the α2 helix is most exposed to the solvent, and in principle avaliable for interactions with additional ligands [180].

In order to establish whether Rpo13 binds to either the major groove or minor groove of DNA or to the DNA phosphate backbone, competition assays were performed with either methyl green and netropsin, which respectively compete for major and minor groove binding. Methyl green and netropsin were found to abolish

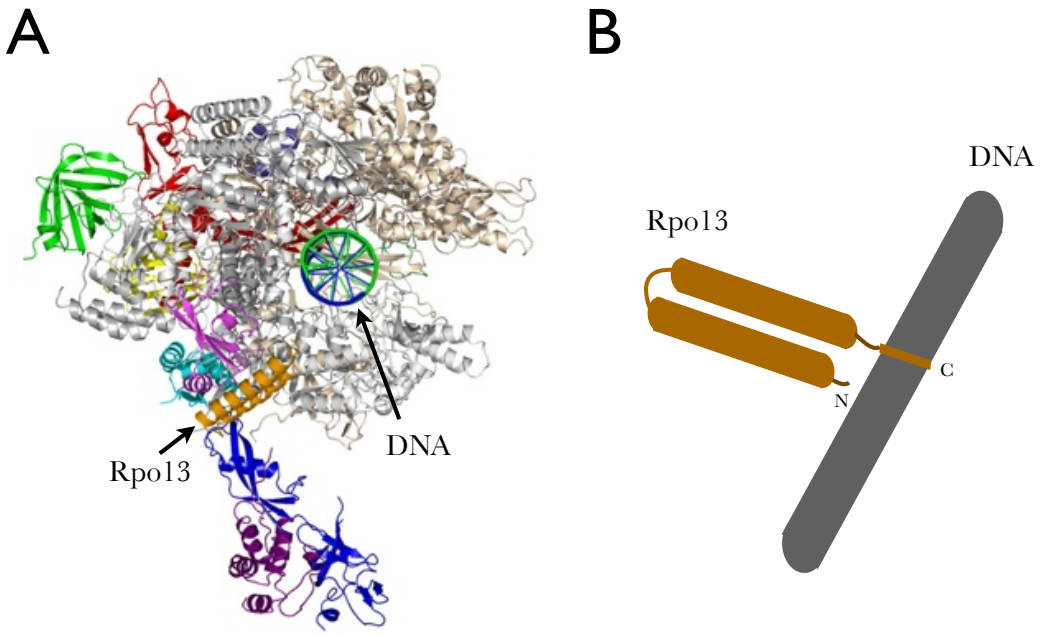


Figure 4.25: *Proposed model for Rpo13's binding to the DNA.* ***(A)****) Crystal structure of the archaeal S. shibatae RNAP in the presence of DNA, where the poorly resolved, flexible C-terminal tail of Rpo13, which is highly positively charged, is seen to point towards the DNA. The PDB code for the archaeal RNAP in complex with the DNA is 4AYB.* ***(B)****) The major determinant of Rpo13's binding to DNA is the highly positively charged tail, which binds to the negatively charged DNA phosphate backbone. Contacts between the flexible N-terminus and the DNA might make small contributions to Rpo13's binding too.*

cooperative TBP and TFB binding to the TATA-box containing oligo, as expected since TBP is known to bind to the TATA-box DNA minor groove while TFB binds to the major groove of the BRE element [110]. On the other hand, no similar effect was observed in the case of Rpo13 upon methyl green and netropsin addition, where only slight inhibition of DNA binding in the presence of netropsin was observed. Therefore, Rpo13 does not bind to either the major or the minor DNA groove but rather through its positively charged C-terminal tail to the negatively charged DNA phosphate backbone.

In order to assess whether binding of Rpo13 to DNA induces torsional DNA stress and supercoiling as implied by the analysis of the archaeal RNAP crystal structure ([180]), a nicked-plasmid based supercoiling assay was used. This assay revealed that, upon increasing of Rpo13 amount, the distribution of topoisomers

shifts from closed circle topoisomers towards the nicked topoisomer. This trend was verified by use of ethidium bromide. These results indicate that inhibition of the ligation reaction occurs upon increasing Rpo13 levels. This is most likely due to the binding of Rpo13 to the nicked plasmid DNA, which in turn prevents the action of the ligase.

A gel filtration procedure was performed to establish whether endogenous Rpo13 within the *S. acidocaldarius* cell extract is all associated with the endogenous, soluble RNAP or whether there is a free Rpo13 pool within the cell extract. While it is known that Rpo13 is a *bona fide* RNAP subunit from a structural point of view [96], the gel filtration procedure would address whether Rpo13 is a true *in vivo* RNAP subunit within the *S. acidocaldarius* cell extract context. Indeed, Rpo13 could act as a transcription factor only temporarily associated with the RNAP, which dissociates from the RNAP once it has finished its transcription factor-like function. It is found that, while the majority of Rpo13 within the *S. acidocaldarius* cell extract is associated with the archaeal RNAP, a small pool elutes at an elution volume corresponding to monomeric recombinant Rpo13. This indicates that not all Rpo13 is associated with the archaeal RNAP within the *S. acidocaldarius* cell extract. Thus, the hypothesis that Rpo13 temporarily associates with the RNAP cannot be excluded. A more rigorous analysis of the *in vivo* distribution of Rpo13 versus the one of bulk RNAP in its elongating form is needed. It is also interesting to note that the RNAP in the *S. acidocaldarius* cell extract has a broader elution profile than the purified RNAP, eluting as a broad peak between the void volume of the column and the same fractions that the purified RNAP occupies. These higher molecular weight complexes may contain additional protein factors in a holo-enzyme complex, such as translation factors, since transcription and translation is thought to be coupled in Archaea [58].

In summary, I have purified the new, archaeal-specific Rpo13 subunit and found it to bind to DNA with low affinity but with specificity for dsDNA. Furthermore, the third, positively charged C-terminal tail is most likely the major determinant of DNA

binding, since I observed no DNA binding activity upon incubation of the ΔC Rpo13 with dsDNA. Moreover, I found that Rpo13 binds to DNA with no specificity for the major or minor DNA groove; rather, it associates with the negatively charged phosphate backbone, and not in such a manner to introduce DNA supercoiling. In addition, within the *S. acidocaldarius* cell extract, not all Rpo13 is associated with the archaeal RNAP, suggesting that Rpo13 might only temporarily associate with the RNAP, having more a transcription factor-like role. Indeed, Rpo13's $\alpha 2$ helix is characterized by polar residues, and it shows some similarity to the charged region of Tfg1-like factors of TFIIF [180]. Finally, I show that in solution Rpo13 forms mainly dimers [179], a behaviour shared with TFIIF, which has a specific dimerisation domain preceding the charged region [61], [39].

To conclude, the archaeal RNAP of the *Sulfolobus* genus is thought to achieve and enhance DNA downstream stabilisation through a hinge movement of the Rpo1 jaw lobe subunit. This might constitutively lock onto the downstream DNA, and by the interaction of a new Rpo13 subunit be able to bind to dsDNA but not ssDNA.

Chapter 5

The Function of Rpo13 in the Context of Transcription

5.1 Introduction

In this study, I propose that Rpo13 interacts with the negatively charged DNA phosphate backbone, with specificity for dsDNA as opposed to ssDNA [180]. Specifically, the third, positively charged C-terminal tail is the major determinant of DNA binding [180]. Therefore, at initiation, the third C-terminal tail could provide a lock point, against which the main RNAP body can push or twist the DNA, while α1 and α2 helices act as constitutive anchors onto the RNAP [96]. During elongation, these locking interactions would be overcome by the translocation forces but may still affect DNA duplex stability [96]. Thus, Rpo13 may perform some of the roles attributed to the eukaryotic-specific initiation factors missing in Archaea.

Indeed, archaeal homologues of TBP and TFB exist and are sufficient for accurate and efficient *in vitro* transcription ([77], [138]). TFE plays only a stimulatory role under conditions where promoter recognition by TBP is sub-optimal ([12], [72]). No archaeal homologues of the remaining transcription factors such as TFIIA and TFIIF have been identified to date. Homologues of the TFIIH subunits XPB and XPD do exist, but are involved in nucleotide excision repair [142], [147].

Furthermore, analysis of reconstituted *in vitro* reactions performed in the presence of ATP or an ATP analogue revealed the same transcription levels ([17]), suggesting that a TFIIH-like ATPase activity is not required. Hence, archaeal promoter opening differs from RNAPII-directed open complex formation in that it is not driven by the helicase action of TFIIH. Therefore, the third C-terminal tail of Rpo13 could act as a barrier against which RNAP can twist the DNA, thus replacing the major role of TFIIH for eukaryotic promoter opening. It would be interesting to remove Rpo13 from the archaeal RNAP in order to directly test the role of Rpo13 in promoter opening.

Moreover, upon initial analysis of the archaeal TBP-TFB-DNA crystal structure, it was thought that the orientation of the TBP-TFB-DNA pre-initiation complex assembly on the promoter is inverted with respect to that observed in the equivalent crystal structures of eukaryotic systems ([97]). Thus, the C-terminal tail of TBP binds to the downstream end of the TATA box as opposed to the upstream end as in the eukaryotic system [97]. Later work however revealed that the polarity of the archaeal TBP-TFB-DNA complex is the same as the eukaryotic TBP-TFIIB-DNA complex, with the BRE promoter element defining such polarity in Archaea [18]. Indeed, an absolute shift in the polarity of transcription was observed in transcription assays using plasmid constructs in which the first 6 or the first 10 bases upstream and downstream of the TATA box were swapped [18]. Also, mutation studies showed that TFB recognises the major groove of DNA upstream of the TATA box, and that residues within the C-terminal repeat of TFB are important for this sequence-specific recognition [18]. This is in contrast to the eukaryotic situation, where transcription polarity is thought to be defined by the intrinsic deformability of the TATA box, where the C-terminal repeat of TBP preferentially contacts the more flexible end of the TATA box ([69]). Alternatively, the highly asymmetric charge potential distribution in TBP ([94]), or the asymmetric positioning of a proline in the C-terminal repeat may be important [87]. Additionally, a number of TAFs in the TFIID complex, which have DNA-binding activities ([164], [31]),

may influence the orientation of TFIID on some promoters through TAF-DNA interactions. Nonetheless, it is of much interest to obtain the crystal structure of the archaeal TBP-TFB-DNA RNAP pre-initiation complex, which would define the mode of interaction of the two transcription factors with the RNAP at the molecular level, and the mode of RNAP's recruitment to the DNA. Indeed, an archaeal homologue of TFIIF, which is involved in helping TFIIB's interaction with the RNAP, is absent [71]. In order to avoid heterogeneity of the pre-initiation RNAP crystal structure due to single binding events of only TBP and TFB as well as due to conformational flexibility during the crystallisation procedure, which would hamper the resolution of the pre-initiation RNAP crystal structure containing both of these transcription factors as opposed to only TBP or TFB, resolution of the crystal structure of the pre-initiation RNAP-DNA complex with a TBP-TFB protein fusion is in current progress (Nicola Abrescia, personal communication). It is nonetheless of interest to test the *in vitro* function of this TBP-TFB fusion with respect to that of cooperative binding of individually purified TBP and TFB. The biological relevance of the TBP-TFB fusion is currently unknown as is the possibility that it is a good representative of TBP's and TFB's binding to the RNAP.

5.2 Results

5.2.1 *In vitro* Function of the TBP-TFB Fusion

In order to test for the *in vitro* function of the TBP-TFB fusion with respect to that of cooperative binding of individually purified TBP and TFB, the TBP-TFB fusion containing five -GSS- repeats as a linker between TBP and TFB (55 kDa) was purified by Magda Wojtas from Nicola Abrescia's laboratory, where it is found that a high salt concentration of 900 mM KCl is needed in order to stabilise the TBP-TFB fusion, which is prone to aggregation at standard and physiologically relevant salt concentrations (Magda Wojtas, personal communication) (Figure5.1).

To test for any DNA binding capability of the TBP-TFB fusion and compare its

DNA binding pattern with respect to that obtained by cooperative binding of TBP and TFB, DNA binding reactions were performed with the dsDNA T6 promoter sequence containing a TATA-like A box [139], [15]. As shown in Figure5.2, the TBP-TFB fusion is capable of DNA binding. Moreover, upon quantitation of DNA binding (Figure5.2), it was found that at lower protein concentrations, the TBP-TFB fusion binds more to the DNA than individually added TBP and TFB proteins at the same molar ratio. On the other hand, at higher protein concentrations, the TBP-TFB fusion binds less to the DNA than individually added TBP and TFB proteins at the same molar ratio (Figure5.2).

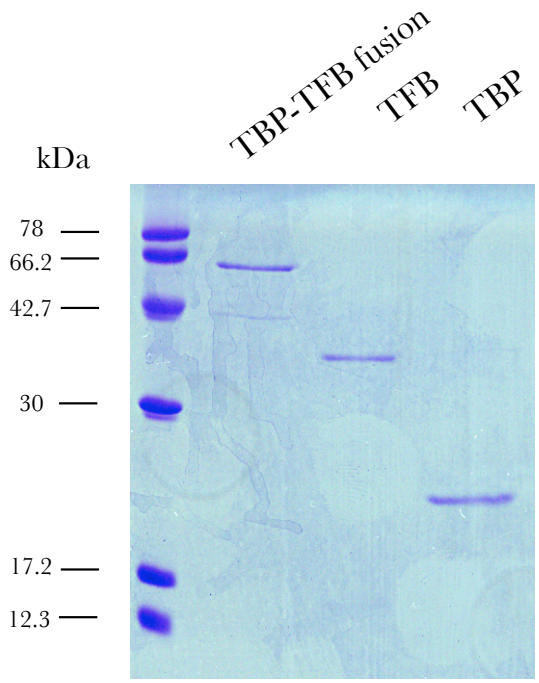


Figure 5.1: *SDS-PAGE gel showing 1.5 µg of TBP, TFB and the TBP-TFB fusion containing five -GSS- repeats as a linker between TBP and TFB. TBP and TFB migrate according to their sizes of 19 kDa and 35 kDa respectively, while the TBP-TFB fusion migrates according to its size of 55 kDa. Staining is achieved with Coomassie Brilliant Blue.*

Furthermore, *in vitro* transcription reactions with a plasmid containing the T6 promoter sequence containing a TATA-like A box ([15]) were performed for both the TBP-TFB fusion and individually added TBP and TFB proteins at the same molar ratio. As shown in (Figure5.3), in addition to DNA binding, the TBP-TFB fusion is

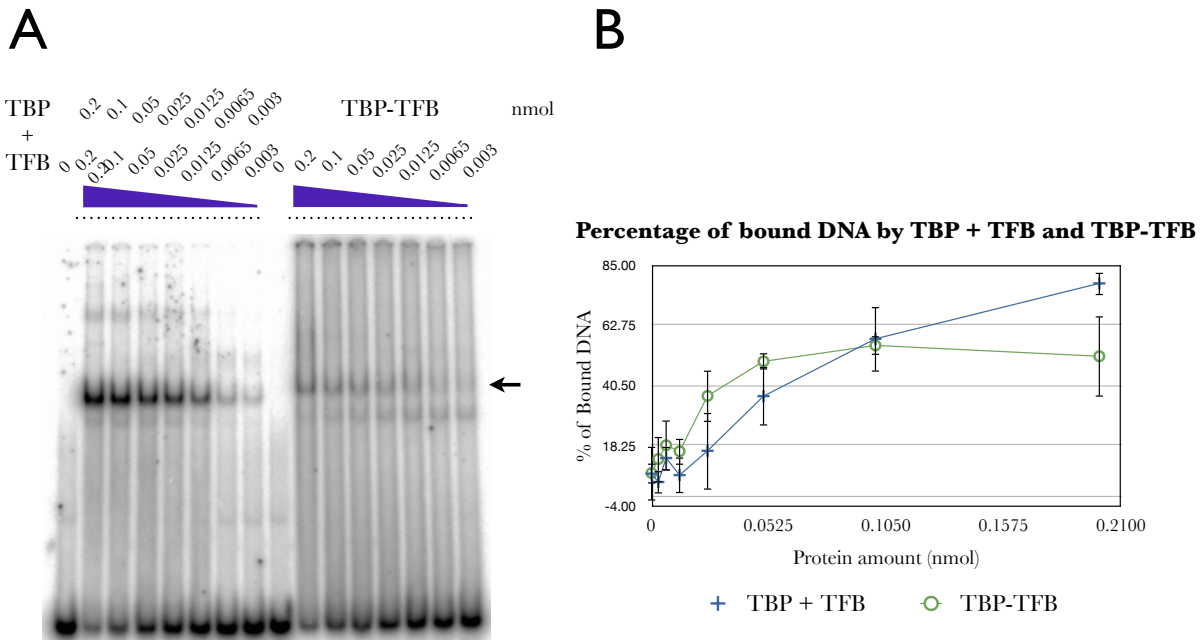


Figure 5.2: *DNA binding patterns of TBP + TFB and TBP-TFB fusion.* ***(A)*** *EMSA assay performed with the T6 promoter sequence containing the TATA-like A box with either TBP + TFB or the TBP-TFB fusion containing five -GSS- repeats as a linker between TBP and TFB. TBP + TFB and the TBP-TFB fusion were added at the same molar ratio. It is found that the TBP-TFB fusion binds to the DNA (arrow) similarly to the cooperative binding of TBP and TFB to the DNA.* ***(B)*** *Upon quantitation of DNA binding, it is found that at lower protein concentrations, the TBP-TFB fusion binds more to the DNA than individually added TBP and TFB proteins. On the other hand, at higher protein concentrations, the TBP-TFB fusion binds less to the DNA than individually added TBP and TFB proteins. All EMSA reactions were performed in triplicate.*

capable to direct transcription, albeit at lower levels than individually added TBP and TFB proteins at the same molar ratio. Background *in vitro* transcription in the absence of factors with only RNAP is also observed, as previously reported [15].

5.2.2 Role of Rpo13 in Transcription

The crystal structure of the archaeal *S. shibatae* RNAP in the presence of bubble DNA oligo obtained by collaborative work performed by Dr. Nicola Abrescia ([180]), showed that the N-terminal Rpo13 region, with the first visible Cα N-terminal Phe33 residue as well as the Glu32 residue, points towards the DNA, and the poorly resolved, flexible C-terminal tail is also in the direction for contact with the DNA. Specifically, as illustrated in (Figure5.4), Rpo13's Glu32 residue contacts

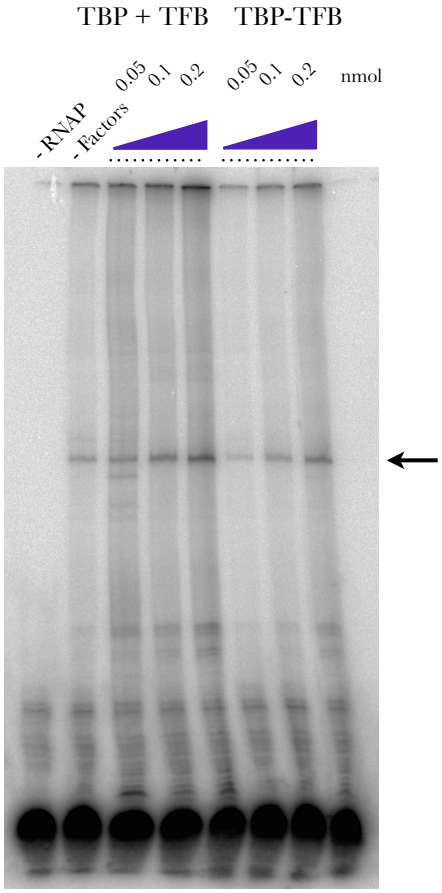


Figure 5.3: *In vitro transcription reactions performed with a plasmid containing the T6 promoter sequence containing a TATA-like A box with either TBP + TFB or the TBP-TFB fusion added at the same molar ratio as well as RNAP. It is found that the TBP-TFB fusion is capable to direct transcription (arrow), albeit at lower levels than individually added TBP and TFB proteins at the same molar ratio. Background in vitro transcription in the presence of only RNAP is also observed, since the RNAP purified from cells is likely to be contaminated with factors. All reactions contain 200 ng of RNAP.*

dsDNA after the DNA bubble, in conjunction with other subunits' residues [180]. I wished to test whether these contacts of Rpo13 with the DNA contribute to DNA unwinding and open complex formation and promote efficient transcription. I therefore performed *in vitro* transcription reactions with purified, endogenous *S. shibatae* RNAP and the bubble oligo substrate, which is 47 nucleotides long, as well as the equivalent DNA sequence in dsDNA linear form. The purpose of this assay is to test whether the RNAP can promote transcription from the bubble oligo substrate, bypassing the requirement of TBP and TFB. As can be seen in Figure5.5,

when the RNAP is incubated with the bubble oligo substrate, a RNA product of the size between 35 and 40 nucleotides is observed, which is present at a lower amount in the reaction containing RNAP and the linear dsDNA, or not present at all. Addition of either TBP of TFB or both of these transcription factors does not change the amount of observed product (Figure5.5). No transcription is observed in the absence of RNAP. Furthermore, the same *in vitro* transcription reactions were performed at different temperatures of 48 °C, 55 °C, 65 °C and 75 °C (Figure5.6). It can be observed that transcription levels of the product of the size between 35 and 40 nucleotides increase with increasing temperature when performed with the bubble oligo substrate, with the most efficient transcription occurring at the physiologically relevant 65 °C temperature (Figure5.6). On the other hand, no difference is observed when the same reactions are performed with the linear dsDNA. No transcription is observed in the absence of RNAP. Thus, incubation of the RNAP with the bubble oligo substrate, which in the RNAP crystal structure is shown to make contacts with the N- and C-termini domains of Rpo13, leads to the formation of a RNA product of the size between 35 and 40 nucleotides specific for the bubble oligo substrate.

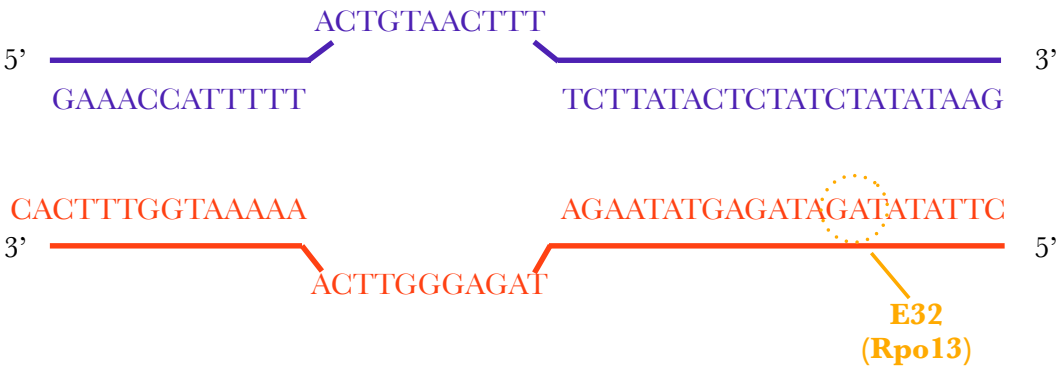


Figure 5.4: *Bubble oligo substrate, which is 47 nucleotides long, used for the crystallisation procedure of the archaeal S. shibatae RNAP in complex with the DNA ([180]). It is found that the Rpo13's Glu32 residue contacts dsDNA after the DNA bubble, in conjunction with other subunits' residues.*

I noted that products bigger than the 47 nucleotide long oligo substrate were synthesised when the *in vitro* transcription reactions were carried on with the oligo in either its bubble form or linear form (Figure5.5), (Figure5.6). In order to test for any

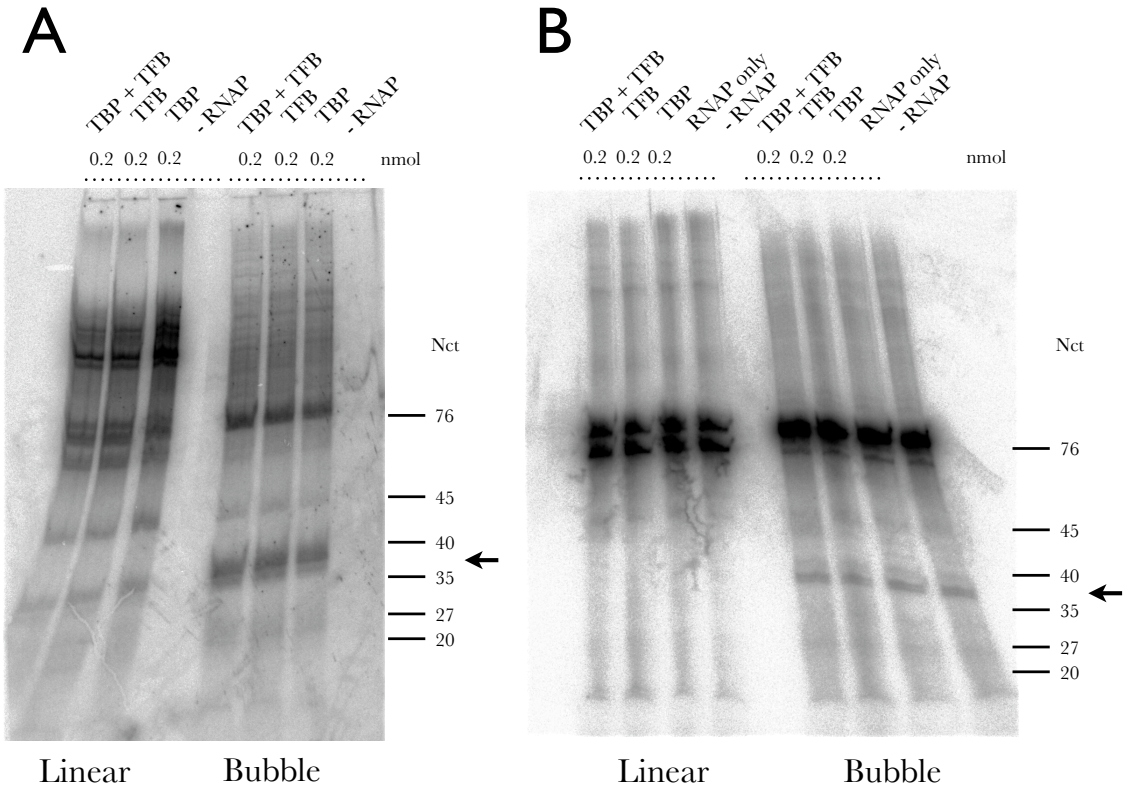


Figure 5.5: *In vitro transcription reactions performed with the bubble oligo substrate in the presence of RNAP alone or in conjunction with the transcription factors.* ***(A)*** *Reactions were performed in the presence of transcription factors, with the DNA oligo either in the bubble form or linear form. A transcription product of a size between 35 and 40 nucleotides is obtained upon incubation with the bubble DNA oligo form (arrow), where the same product is obtained at a lower amount when reactions were performed with the linear DNA oligo form.* ***(B)*** *Reactions were performed either in the presence of transcription factors, or RNAP alone, with the DNA oligo either in the bubble form or linear form. No difference in the obtained products is observed upon the addition of transcription factors. A transcription product of a size between 35 and 40 nucleotides is obtained only upon incubation with the bubble DNA oligo form (arrow). All reactions contain 200 ng of RNAP.*

competing primer-extension reaction, where hybrid DNA-RNA products are formed which could account for these larger products, *in vitro* transcription reactions were performed followed by DNA digestion with DNase, which would digest any DNA and leave only the RNA products. As shown in Figure5.7, upon DNase digestion, the product's size is reduced to products smaller than 40 nucleotides in both the bubble oligo substrate and linear dsDNA. This indicates the presence of DNA-RNA hybrid products in the absence of DNase digestion of size bigger than 47 nucleotides, which are formed upon RNAP's *de novo* RNA synthesis with priming on the DNA

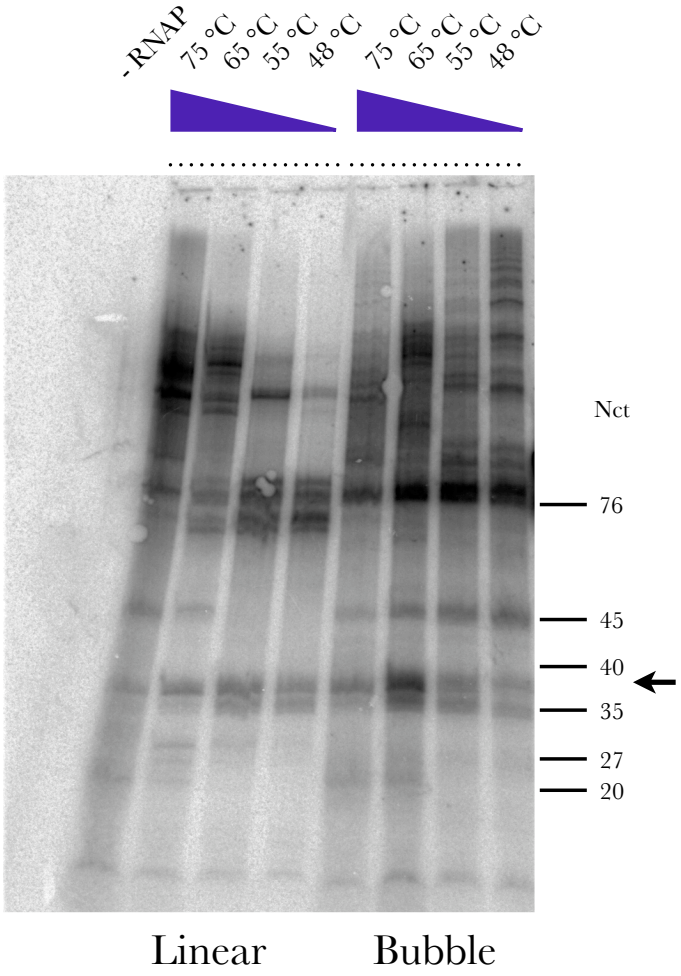


Figure 5.6: *In vitro transcription reactions performed with the bubble oligo substrate in the presence of RNAP alone and at different temperatures, ranging from 48 °C to 75 °C. An increase in the amount of transcribed product of the size between 35 and 40 nucleotides is observed upon increasing temperature only for the bubble DNA oligo form (arrow), with the most efficient transcription occurring at the physiologically relevant 65 °C temperature. All reactions contain 200 ng of RNAP.*

bubble oligo substrate and linear dsDNA.

5.2.3 Attempts at Removing Rpo13 from the Archaeal RNAP

To test for a more direct role of Rpo13 in RNAP *in vitro* transcription by promoting DNA unwinding and open complex formation, a removal procedure based on the one described by Lanzerdorfer and coworkers was performed [104]. Indeed, obtaining a

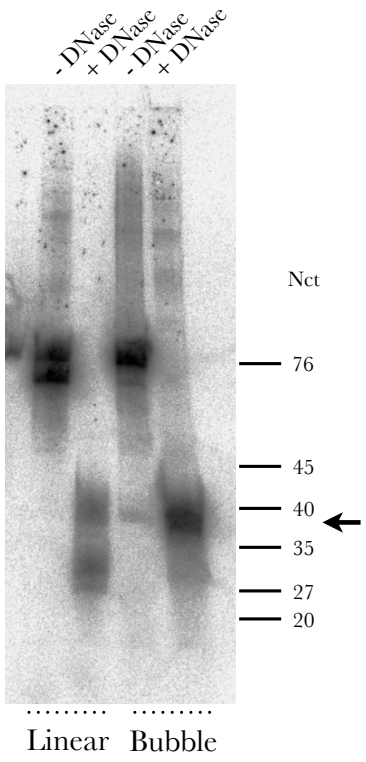


Figure 5.7: *Products bigger than the 47 nucleotide long oligo substrate are observed when the in vitro transcription reactions are carried on either with the oligo in its bubble form or linear form. Upon DNase digestion, the product's size is reduced to products smaller than 40 nucleotides. This indicates the presence of DNA-RNA hybrid products in the absence of DNase digestion of size bigger than 47 nucleotides, possibly due to a competing primer-extension reaction carried on by the RNAP.*

RNAP enzyme deficient in Rpo13 would allow comparison of its ability to direct *in vitro* transcription with wild-type RNAP, and so directly assess the role of Rpo13 in transcription. According to the procedure described by Lanzerdorfer and coworkers, an ammonium acetate gradient resolves the RNAP bound to a positively charged MonoQ column into different RNAP peaks, the first one being deficient in the Rpo13 subunit and the remaining ones being RNAP variants each containing a different Rpo13 component [104]. I followed this procedure (Figure5.8), and obtained several RNAP peaks as previously described [104]. Each peak was characterised in terms of Rpo2 (120 kDa) and Rpo13 (20 kDa) subunits, by western blots performed with corresponding antibodies. However, the very first peak, previously identified as an RNAP deficient in the Rpo13 subunit (elutes with an ammonium acetate percentage

of around 50%) was found to contain Rpo13, albeit of a different size from Rpo13 in the input RNAP sample (Figure5.8). The other RNAP peaks also display different sized Rpo13s. Fractions corresponding to the first peak were pooled together, and subjected to another ammonium acetate gradient in a further attempt to remove Rpo13. However, as can be seen (Figure5.8), fractions corresponding to the RNAP peak are still found to contain Rpo13 in a different form from to that of Rpo13 in the input sample. Furthermore, the binding capacity of Rpo13 to the MonoQ column was tested (Figure5.9), as this underpins the removal of Rpo13 from RNAP through the ammonium acetate gradient. The affinity of Rpo13 for the MonoQ column should be different to that of RNAP. As can be seen (Figure5.9), *S. acidocaldarius* Rpo13 is found to bind to the MonoQ column, as well as *S. shibatae* Rpo13 (data not shown). Furthermore, it elutes with a higher percentage of ammonium acetate of around 40%, which indicates a different and lower Rpo13 affinity for the MonoQ column than the first RNAP peak.

As an alternative strategy, an experiment was set up, with excess recombinant His-tagged Rpo13 incubated with recombinant, purified RNAP. If subunit exchange occurs between soluble, recombinant Rpo13 and the endogenous Rpo13 in RNAP, then upon passing of the incubation samples through a Ni^{2+}-NTA-column, His-tagged Rpo13 should pull-down the entire RNAP. Both incubation reactions were performed with buffer containing either 1% of Triton or 1 M of NaCl. As expected (Figure5.10), upon incubation of buffer control only with no Rpo13, all RNAP was found in the supernatant as monitored by western blot performed with the antibody against Rpo2. Upon incubation of either 1% Triton or 1 M NaCl buffer containing Rpo13, all RNAP was also found in the supernatant as monitored with a western blot. Therefore, binding of His-tagged Rpo13 to the Ni^{2+}-NTA beads does not pull down the entire RNAP.

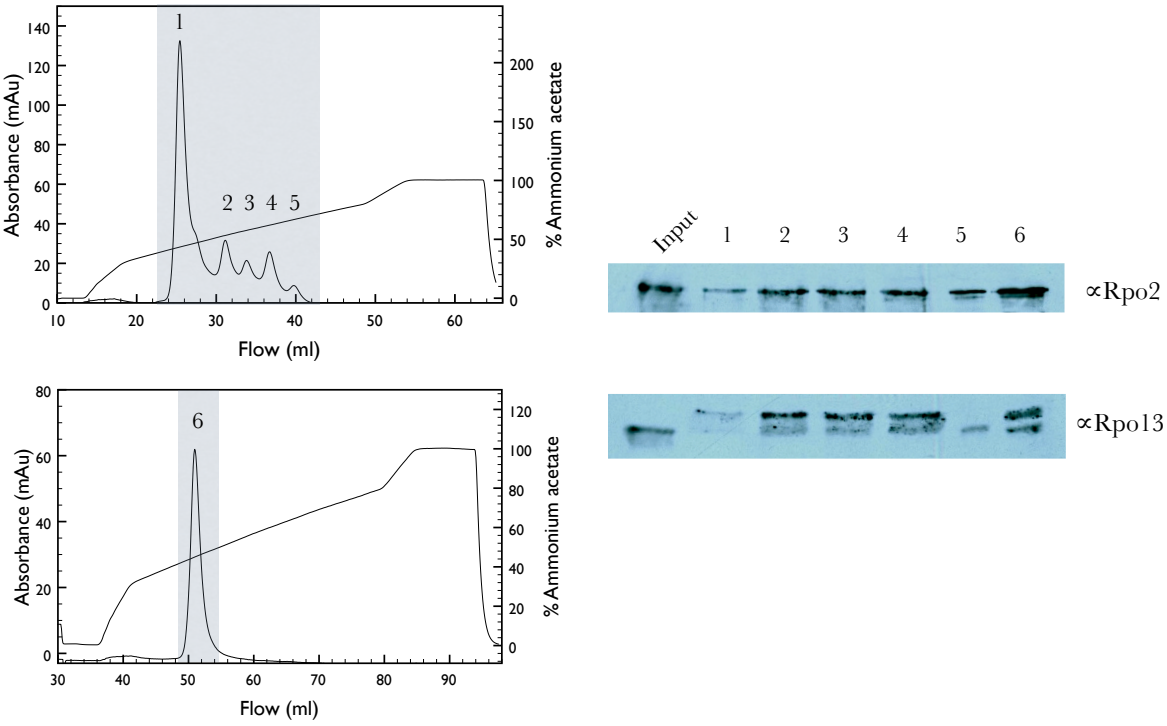


Figure 5.8: *Attempt at resolving different RNAP forms with an ammonium acetate gradient. Five different RNAP peaks (1-5) are found to elute upon the application of the ammonium acetate gradient. The first RNAP peak, previously identified as deficient in the Rpo13 subunit, elutes with an ammonium acetate percentage of around 50%. However, upon western analysis with antibodies against Rpo2 and Rpo13, it is found that the first RNAP peak contains Rpo13, albeit of a different form from that of the Rpo13 of the input RNAP sample. Repetition of the ammonium acetate gradient on the first RNAP peak (1) leads to another RNAP peak (6), which is still found to contain Rpo13 of a different form from that of the Rpo13 of the input RNAP sample.*

5.3 Discussion

The TBP-TFB fusion containing five -GSS- repeats as a linker between TBP and TFB is found to need 900 mM KCl to be stabilised in solution, since it is prone to aggregation at standard and physiologically relevant salt concentrations. This is most likely due to protein aggregation through hydrophobic interactions. A high salt concentrations, ions coat the molecule instead of water molecules, leaving more free water molecules for protein solvation. Furthermore, the TBP-TFB fusion is found to not only bind the dsDNA T6 promoter sequence containing a TATA-like A box similarly to cooperative TBP and TFB binding, but also shows a different

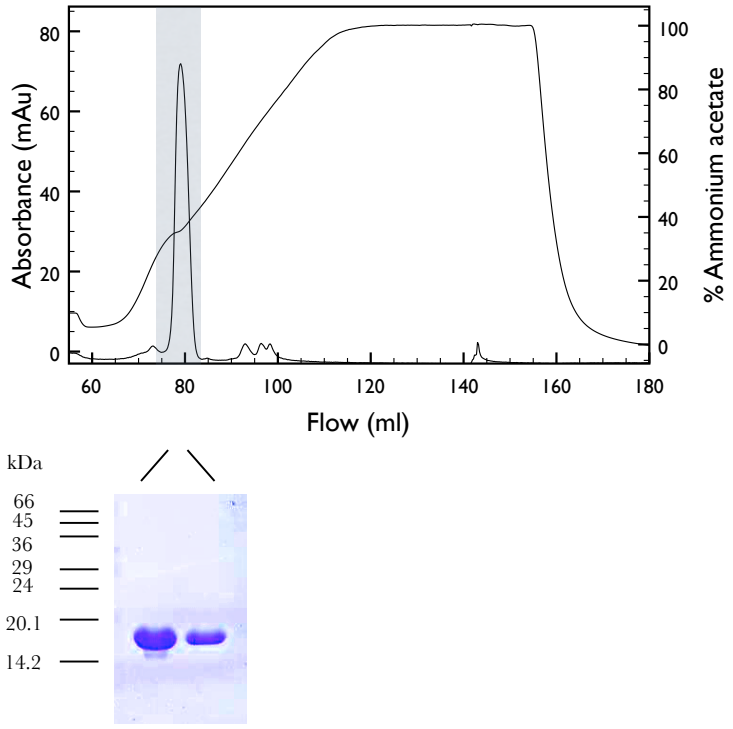


Figure 5.9: *Recombinant S. acidocaldarius Rpo13 is found to bind to the MonoQ column, where it elutes with a lower percentage of ammonium acetate of around 40%. This indicates a lower Rpo13's affinity for the MonoQ column than the first RNAP peak obtained upon the application of an ammonium acetate gradient on the RNAP sample.*

binding pattern at various protein concentrations to that obtained from cooperative TBP and TFB binding. Specifically, at lower protein concentrations, the TBP-TFB fusion binds more to the DNA than individually added TBP and TFB proteins at the same molar ratio. On the other hand, at higher protein concentrations, the TBP-TFB fusion binds DNA less than individually added TBP and TFB proteins. This can be explained by the fact that, at lower protein concentrations, simultaneous binding events of both TBP and TFB fused together are more likely than unique and separate binding events of separate TBP and TFB. However, upon increasing protein concentrations, a cooperative effect between TBP and TFB occurs, where not only TBP binding to DNA recruits TFB, but also TFB interactions with the TATA-box in addition to the BRE element favours TBP binding. It is possible that the geometry of the five -GSS- repeats linker constrains the most favourable

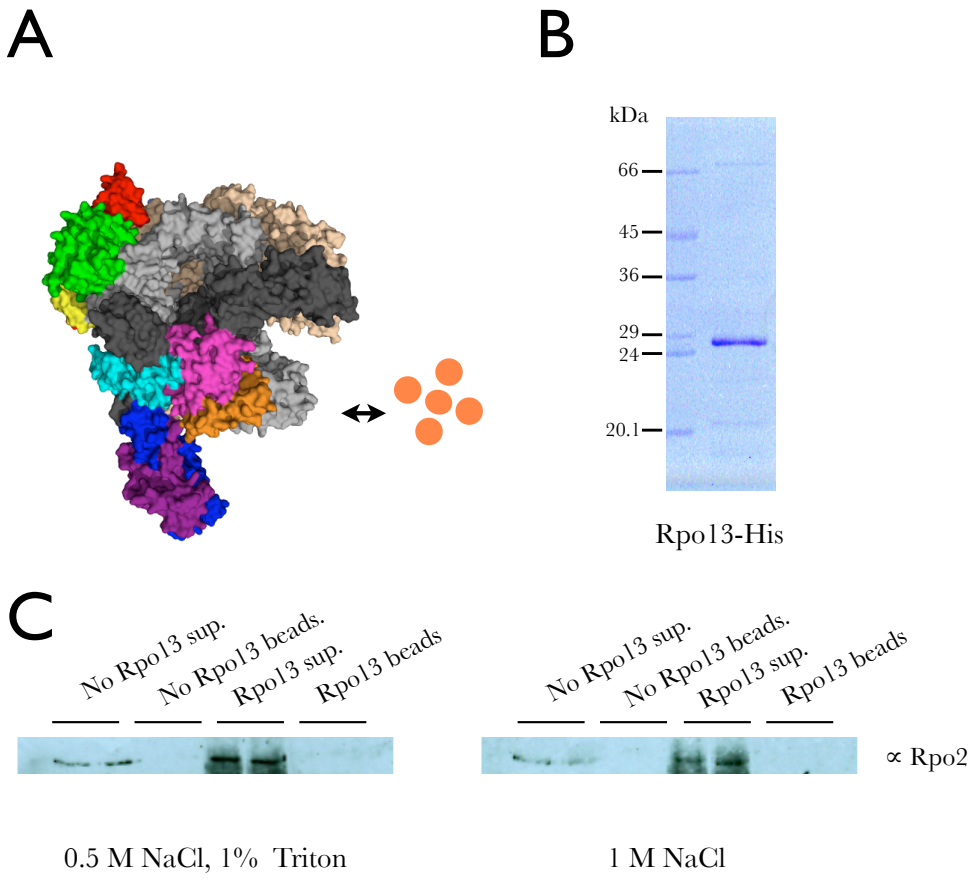


Figure 5.10: *Incubation procedure* ***(A)****, where recombinant, archaeal RNAP is incubated with an excess of recombinant, His-tagged Rpo13* ***(B)****. Upon incubation of buffer control only with no Rpo13 and passing through the Ni^{2+}-NTA beads, all RNAP is found in the supernatant as monitored with a western blot performed with the antibody against Rpo2 for both buffers containing either 1% of Triton or 1 M of NaCl. Upon incubation of buffer containing Rpo13 and passing through the Ni^{2+}-NTA beads, all RNAP is also found in the supernatant as monitored with a western blot performed with the antibody against Rpo2 in both buffer conditions. Therefore, the RNAP is not pulled down upon binding of His-tagged Rpo13 to the Ni^{2+}-NTA beads.*

TBP and TFB binding modes to DNA, preventing contacts between TBP and TFB which underpin the cooperative effect. Most importantly, the TBP-TFB fusion is found to be able to direct *in vitro* transcription of RNAP, albeit at lower levels, most likely due to the lower binding affinity to DNA. Thus, the TBP-TFB fusion is physiologically relevant. A crystal structure of the RNAP-DNA pre-initiation complex together with TBP-TFB fusion would yield important insights into the mode of RNAP recruitment to DNA by the transcription factors.

Archaea lack homologues of the remaining eukaryotic transcription factors involved in pre-initiation complex formation and promoter escape. The binding of Rpo13 to DNA has been proposed to have an effect in transcription initiation, where the third C-terminal tail of Rpo13 could act as a barrier against which RNAP can twist the DNA, thus replacing the major role of TFIIH in eukaryotic promoter opening [96]. Upon analysis of the crystal structure of the archaeal *S. shibatae* RNAP in the presence of bubble DNA oligo obtained by collaborative work performed by Dr. Nicola Abrescia ([180]), it was found that the N-terminal Rpo13 region, with the first visible Cα N-terminal Phe33 residue as well as the Glu32 residue, points towards the DNA, and the poorly resolved, flexible C-terminal tail is also in the direction for contact with the DNA. Since the N-terminal contacts with the DNA are made after the DNA bubble, it is thought that, in conjunction with other subunits residues, Rpo13 contacts with the DNA contribute to DNA unwinding by pulling apart the DNA strands. The *in vitro* transcription assays performed with the RNAP and the bubble oligo substrate as well as the same DNA sequence in dsDNA linear form reveal a product of around 35-40 nucleotides only in the bubble oligo substrate case. No product of 22 nucleotides is observed, which would arise from the internal TSS. Thus, initial localised DNA melting is not enough for the RNAP's subunits including Rpo13 to make the remaining contacts along the DNA oligo and contribute to DNA strand pulling for efficient transcription. Instead, the product of around 35-40 nucleotides most likely arises from terminal transcription initiation due to the breathing of the bubble oligo substrate at its ends due to the high temperature and due to the overhang at the 3' end. Furthermore, products of a higher size than 47 nucleotides appear to be due to either DNA contamination from the RNAP preparation or a RNAP's competing primer-extension reaction. The size of these products is indeed reduced upon DNase digestion, indicating the presence of DNA-RNA hybrid molecules. It may be possible that the oligo DNA forms DNA hairpins, upon which the RNAP extends the DNA ends in *de novo* RNA synthesis, as previously reported for the T7 RNAP and the viral RNAPs from the

poliovirus and the SARS-coronavirus [148], [125], [144], [157]. In summary, my *in vitro* transcription assays argue against the idea that Rpo13 makes valuable DNA contacts further down the DNA template bubble, favouring DNA unwinding and thus transcription. Rpo13's direct role in transcription needs to be assessed by its removal from the RNAP. The activity of the Rpo13-decifient RNAP could then be compared to wild-type RNAP. This would allow a more direct appreciation of Rpo13 function.

A removal procedure based on Lanzerdorfer and coworkers [104] was performed, using an ammonium acetate gradient to resolve RNAP binding to a positively charged MonoQ column. This should yield different RNAP peaks, the first one being deficient in the Rpo13 subunit. This method is based on the idea that Rpo13 is a labile subunit, and its binding affinity for the MonoQ within the RNAP context is higher than the binding affinity for the remaining RNAP. Hence, it should bind more strongly to the column than the rest of the RNAP. A higher ammonium acetate percentage is needed for its elution, while a lower one is needed for RNAP elution, allowing for the separation of the two protein components. By running recombinant Rpo13 on the MonoQ column, I found that it binds to the MonoQ column, eluting at around 40% ammonium acetate, as opposed to a percentage of around 50% for elution of RNAP from *S. acidocaldarius* cell extract. Indeed, the negatively charged N-terminal region of Rpo13 mediates its binding to the positively charged MonoQ matrix. However, when the first peak corresponding to purified RNAP as well as the remaining RNAP peaks were characterised in terms of Rpo2 and Rpo13 subunits with western blots performed with their corresponding antibodies, all were found to have the Rpo13 subunit, albeit of a different form from Rpo13 in the input RNAP. This is most likely explained by different charge densities due to different methylation patterns of lysine residues [23], which will create different Rpo13 mobilities. It is possible that Rpo13 within the RNAP context does not bind the MonoQ column as well as recombinant Rpo13. Also, Rpo13 methylation could shield Rpo13 affinity for the MonoQ column. Possibly purified RNAP obtained from the *S. acidocaldarius*

culture used in this study has a different overall methylation pattern for the different subunits including Rpo13 than that employed by Lanzerdorfer and coworkers ([104]). Furthermore, the various observed RNAP peaks could be due to the presence of DNA contaminant that also binds to the positively charged MonoQ matrix.

Finally, an alternative procedure consisting in incubating His-tagged Rpo13 with purified RNAP was set up to promote subunit exchange between soluble, recombinant Rpo13 and the endogenous Rpo13. If Rpo13 subunit exchange occurs between recombinant Rpo13 and endogenous RNAP, then by passing the incubation samples through a Ni^{2+}-NTA-column, His-tagged Rpo13 should act to pull-down the entire RNAP. However, no pull down of RNAP by the Rpo13 subunit was observed, indicating that no exchange of subunits occurred in the two buffers used. Indeed, 1% of Triton as well as 1 M of NaCl were insufficient to disrupt hydrophobic and electrostatic interactions between the endogenous Rpo13 and the RNAP respectively and so promote subunit exchange.

In summary, I showed that TBP-TFB fusion containing five -GSS- repeats as a linker between TBP and TFB is functional, in that it not only binds to the DNA similarly to cooperative binding of TBP and TFB, but it is also able to direct transcription. This is a complementary analysis to the structural studies currently performed by collaborative work by Dr. Nicola Abrescia, which will yield insights into the crystal structure of the archaeal RNAP-TBP-TFB complex. The crystal structure will enable the understanding of the mode of RNAP recruitment to the DNA by TBP and TFB in Archaea. Furthermore, no archaeal homologues of the eukaryotic transcription factors that play pivotal roles in eukaryotic transcription initiation are currently identified. I propose that the newly identified, archaeal-specific Rpo13 subunit performs some of the roles attributed to the eukaryotic-specific initiation factors missing in Archaea. It has been shown that both the Rpo13 N-terminus and flexible C-terminal tail make contact with a bubble DNA oligo substrate in the new archaeal RNAP-DNA crystal structure [180]. It is thought that the contacts of Rpo13 with DNA help unwind it, so promoting

transcription, in analogy to the role played by eukaryotic transcription factors. Whilst no transcription product arising from the internal TSS of the bubble DNA oligo substrate is observed, the removal of Rpo13 from RNAP by alternative methods than the ones described in this chapter may allow a more direct role of Rpo13 in transcription to be appreciated.

Chapter 6

Interaction between RPC34 and the Single Archaeal RNAP

6.1 Introduction

Recent analysis of archaeal genomes has identified an archaeal homologue of the eukaryotic RNAPIII-specific RPC34 subunit [22], which together with RPC82 and RPC31 forms the RNAP-III-associated C82/34/31 complex [56]. The archaeal RPC34 was predicted to have a winged-helix (WH) domain as well as a zinc-finger domain shared with most eukaryotic RPC34 sequences (Figure6.1) [22]. Thus, Archaea not only contain the TFE factor, which is monomeric and corresponds to the N-terminal half of the TFIIE-α subunit involved in RNAP-II modulation, but also RPC34, the eukaryotic homologue involved in RNAP-III transcription. This later protein is related to the TFIIE-β subunit, in that it contains two adjacent WH domains in the N-terminal region similarly to TFIIE-β [61].

Analysis of the genomic context of the archaeal RPC34 revealed the co-localisation of this gene with factors involved in the transcription, modification and processing of structural rRNAs and tRNAs, such as a tRNA pyrophosphorylase, and a paralog of TFB referred to as TFB-2 (Figure6.2) [22]. Therefore, TFB-2 may interact with the archaeal RPC34 to promote transcription of structural

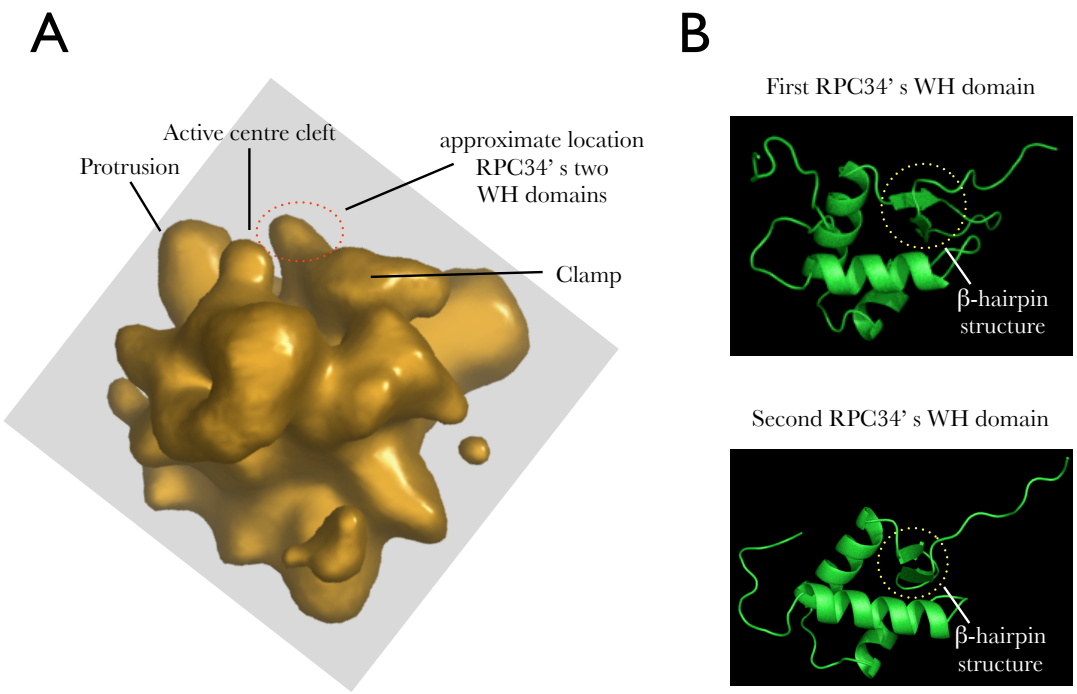


Figure 6.1: *RPC34's location on the eukaryotic RNAP-III.* ***(A)*** *Location of the eukaryotic RPC34 on RNAP-III, where the two RPC34's WH domains* ***(B))*** *are assigned to a prominent density which forms a suspension over the cleft from the clamp to the protrusion. Indeed, the two WH domains (PDB codes 2DK5 and 2DK8) fit the two lobes of the structure, and RPC34 crosslinks to promoter DNA around position +21 [11], which is adjacent in the homologous Pol II promoter complex model [98].* ***(B)*** *The WH domains are characterised by a two-strand β sheet which forms a prominent β shairpin structure (wing) and a bundle of α helices. The figure is recreated from [161], where the electron microscopy structure of the RNAP-III is EMD-1753.*

promoters similarly to the eukaryotic interaction between RPC34 and TFIIIB70 [29], [91], [173]. Indeed, while TBP and TFIIB are necessary and sufficient for the RNAP-II recruitment to the DNA, RNAP-III requires RPC34 to mediate the interaction between TFIIIB70 and RNAP-III[29], [91], [173]. Therefore, the archaeal RPC34-TFB-2 complex may associate with the single archaeal RNAP similarly to the association of RPC34-TFIIIB70's with the RNAP-III.

Qureshi and coworkers showed that transcription from the structural 16S/23S rRNA promoter could be successfully initiated in the absence of RPC34. This indicates that there is no strict RPC34 requirement for this promoter recognition and RNAP recruitment. The different transcription levels obtained by the reconstituted system and crude extract suggest potential regulatory roles for further transcription

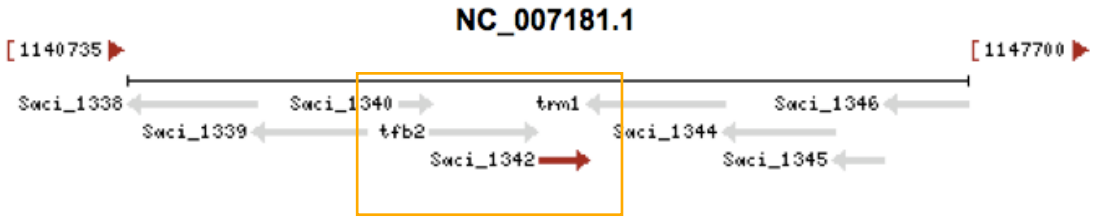


Figure 6.2: *The S. acidocaldarius RPC34 gene (Saci_1342) is colocalised with factors involved in the transcription, modification and processing of structural rRNAs and tRNAs. Examples include TFB-2, which is a TFB homologue, Saci_1340, which functions in a trimeric complex to guide RNA-target RNA complexes for the purposes of pseudouridylation, Saci_1343, which is a N(2),N(2)-dimethylguanosine tRNA methyltransferase, and Saci_1346, which is a fibrillarin-like rRNA methylase.*

factors absent in the reconstituted system, which may include RPC34 [138]. The WH domain of RPC34 might mediate protein-DNA interactions, considering that eukaryotic RPC34 was cross-linked to DNA in transcription initiation complexes, where it is located at the most upstream position relative to the start site [11]. Indeed, the eukaryotic RPC34 subunit is located to the RNAP-III saddle and stalk [56], where its WH domains are located over the cleft from the clamp to the protrusion similarly to TFIIE, and contact the DNA (Figure6.1)[161]. Furthermore, RPC34 was also found to influence the formation of the open promoter complex, since the initiation reaction of a RNAP-III mutant with the mutagenised RPC34 subunit, deficient in TFIIIB70 recruitment, was found to be temperature dependent [29]. This indicates that a proper RPC34-TFIIIB70 interaction is not only necessary for correct RNAP-IIII recruitment, but also for open complex formation [29]. Alternatively, RPC34 might simply help the recruitment of the archaeal RNAP through interaction with its clamp domain and, similarly to the RNAP-III system [136], increase transcription levels. It is therefore of much interest to analyse both the *in vitro* and *in vivo* RPC34 interaction with the single, archaeal RNAP, in conjunction with RPC34's interaction with TFB-2. If such interactions exist, the association of the complex formed by RPC34 and TFB-2 with the RNAP would define a new form of archaeal RNAP. Furthermore, the complex could modulate the

specificity of the archaeal RNAP for the transcription of a subset of genes, such as non-coding rRNA and tRNA genes, in analogy to the RNAP-III/TFIIIB70/RPC34 system. Thus, the split of RNAP into the various, gene-class specific forms may have already occurred in Archaea.

6.2 Results

6.2.1 *In vitro* RPC34's Interaction with the Archaeal RNAP

In order to test for an in vitro interaction between RPC34 and the single archaeal RNAP, a glutathione-S-transferase (GST)-pull down assay was performed. In this assay, either GST protein on its own or a RPC34-GST fusion was incubated with either buffer control or *S. acidocaldarius* cell extract. If the RNAP binds to RPC34, a band corresponding to the Rpo2 subunit (120 kDa) should be seen by western blot analysis with the antibody against the Rpo2 subunit. As shown in (Figure6.3), upon western blot analysis of the samples, a band corresponding to the size of the Rpo2 subunit of 120 kDa is seen only when the RPC34-GST fusion is incubated with the *S. acidocaldarius* cell extract but not with the buffer control. No band was detected when the GST protein is incubated with either the buffer control or *S. acidocaldarius* cell extract.

In order to analyse RPC34's interaction with the various, individual RNAP subunits, a yeast two-hybrid screen was subsequently performed. In this screen, the various RNAP subunits are encoded in fusion with the Gal4 activation domain, while RPC34 is encoded in fusion with the Gal4 binding domain. If interaction of RPC34 occurs with any of the RNAP subunits, yeast growth occurs in the selective *-his* plate. The *his* reporter gene in the AH109 yeast strain is under the control of the Gal4-responsive promoter element, which is activated upon binding of both Gal4 activation and binding domains brought together by the interacting partners. As can be seen in Figure6.4, all yeast strains grow in the permissive *-trp* and *-leu* condition, which select for both coding plasmids. Furthermore, selective yeast

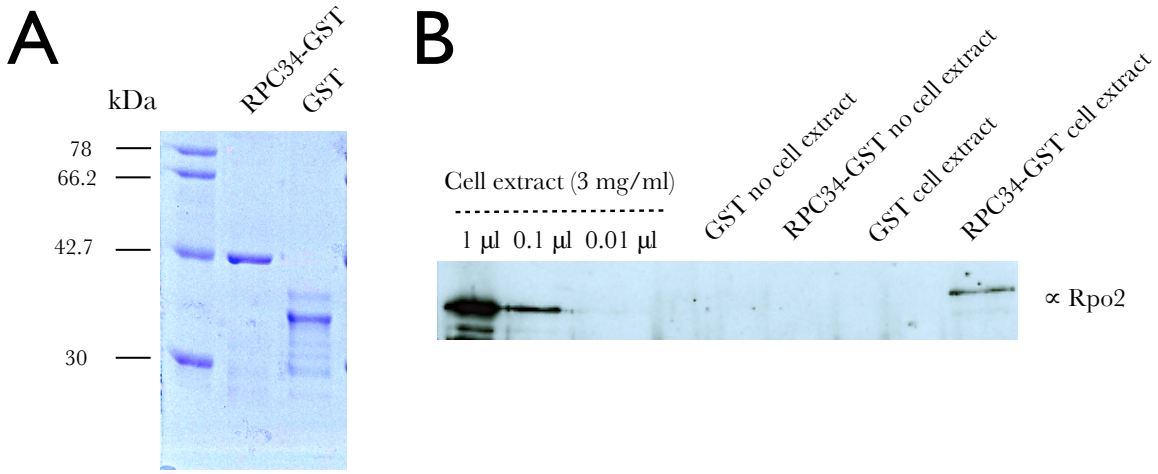


Figure 6.3: *A GST-pull down assay, where both GST and RPC34-GST (left) are incubated with either buffer control or S. acidocaldarius cell extract (3 mg/ml). Upon western blot analysis of the samples, a band corresponding to the size of the Rpo2 subunit of 120 kDa is seen only when the RPC34-GST fusion is incubated with the S. acidocaldarius cell extract but not with the buffer control, or when the GST protein on its own is incubated with either the buffer control or S. acidocaldarius cell extract.*

growth in the *-trp*, *-leu* and *-his* condition is observed upon yeast co-transformation with the plasmids encoding RPC34 in fusion with the Gal4 binding domain and the Rpo8 subunit in fusion with the Gal4 activation domain. As a positive control yeast growth was also obtained with the previously reported interaction between Gins15 and RecJ components of the DNA replication machinery (Figure6.4) [118]. No yeast growth is observed upon yeast co-transformation with empty plasmids. The equivalent selective yeast growth is not obtained when the yeast is co-transformed with both empty plasmid coding only for the Gal4 binding domain and the Rpo8 subunit in fusion with the Gal4 activation domain (Figure6.4). However, selective growth is observed when the yeast is co-transformed with empty plasmid coding only for the Gal4 binding domain and either the Rpo3, Rpo4, Rpo6, Rpo7 and Rpo12 subunits in fusion with the Gal4 activation domain (Figure6.4). Finally, as shown in (Figure6.4), all yeast strains show no growth in the presence of 2.5 mM 3–amino–1,2,4–triazole (3–AT), which is an inhibitor of the product of the *his3* gene, imidazoleglycerol-phosphate dehydratase, involved in histidine synthesis, except for

yeast growth obtained upon interaction of Gins15 with RecJ.

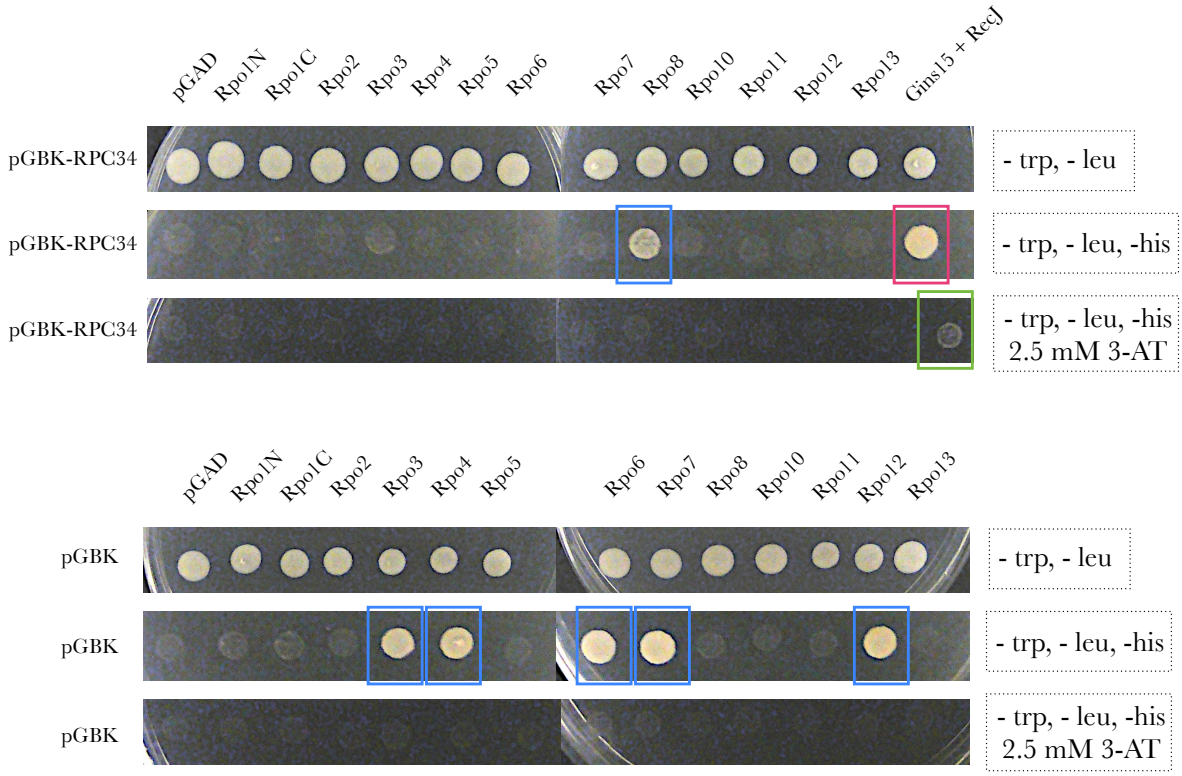


Figure 6.4: *Yeast two-hybrid screen, where an interaction between RPC34 and the RNAP's subunits is tested. Yeast strains co-transformed with plasmids coding for either pGBK-RPC34 or pGBK in conjunction with the individual RNAP subunits encoded in pGAD plasmids grow the in the permissive -trp and -leu condition, which selects for pGBK and pGAD plasmids. However, upon growth in the selective -trp, -leu and -his condition, only the yeast strains co-transformed with plasmids coding for pGBK-RPC34 and the Rpo8 subunit encoded in the pGAD plasmid (blue box) as well as plasmids coding for Gins15 and RecJ (pink box), used as a positive control, are observed to grow. Furthermore, upon growth in the selective -trp, -leu and -his condition, only the yeast strains co-transformed with plasmids coding for pGBK and the Rpo3, Rpo4, Rpo6, Rpo7 and Rpo12 subunits encoded in the pGAD plasmid (blue boxes) are observed to grow. Finally, all yeast strains do not show any growth in the presence of 2.5 mM 3-AT, except for yeast growth obtained upon interaction of Gins15 with RecJ (green box).*

6.2.2 RPC34's Cloning and Purification

In order to study the role of RPC34 from an *in vivo* perspective, a plasmid coding for *S. acidocaldarius* His-tagged RPC34 was generated as described in the Chapter 2 (Materials and Methods). RPC34 was found to be insoluble and in the pellet fraction after cell disruption, indicating that RPC34 is expressed in inclusion bodies

(Figure6.5). Therefore, RPC34 was subsequently purified in its denatured form in the presence of 8 M urea. As can be observed in Figure6.5, an enriched preparation of RPC34 was obtained (molecular weight of 15 kDa).

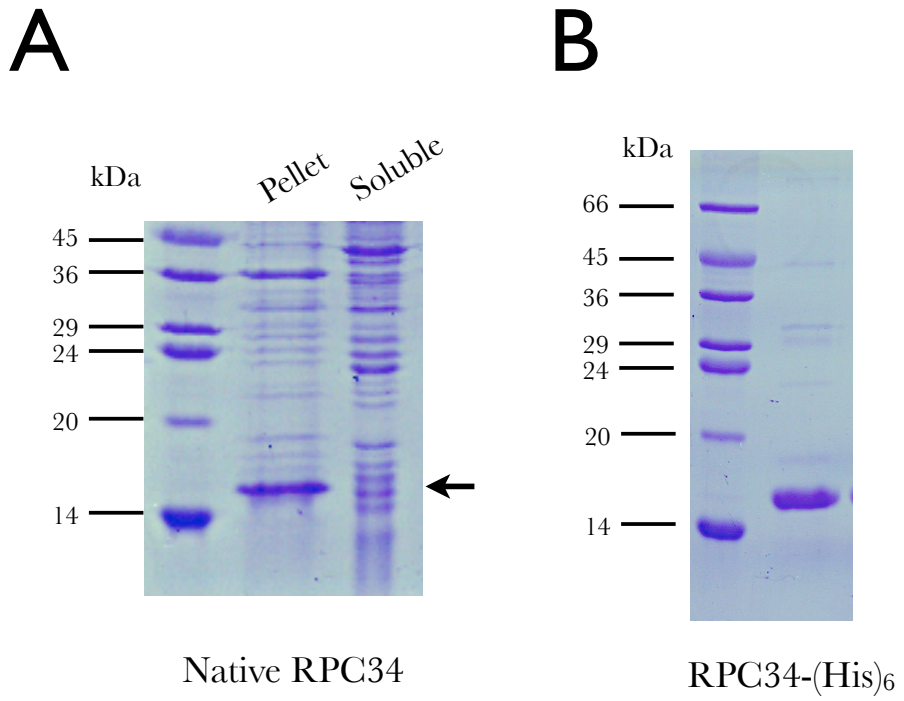


Figure 6.5: *RPC34 expressed under standard conditions is found to be expressed in the inclusion bodies, and hence in the pellet fraction of the cell lysate after centrifugation* ***(A)****). RPC34 expressed in the presence of 8 M urea is found to be soluble, and in its denatured form, where it is found to migrate in the SDS-PAGE gel subsequently stained with Coomassie Brilliant Blue according to its molecular weight of 15 kDa* ***(B)***

.

6.2.3 Antibody against RPC34

A rabbit antibody was raised against the purified, recombinant and denatured *S. acidocaldarius* RPC34. As shown in Figure6.6, upon incubation of the pre-immune serum control with *S. acidocaldarius* cell extract and pure RPC34, no bands corresponding to the size of RPC34 are observed. On the other hand, upon incubation of the antibody from the final-bleed serum with *S. acidocaldarius* cell extract and pure RPC34, similar bands corresponding to the size of RPC34 of 15 kDa were observed (Figure6.6), indicating a successful rabbit immune-response against

recombinant and denatured RPC34. Most importantly, the antibody raised against recombinant, denatured RPC34 also recognises the endogenous RPC34 in the *S. acidocaldarius* cell extract.

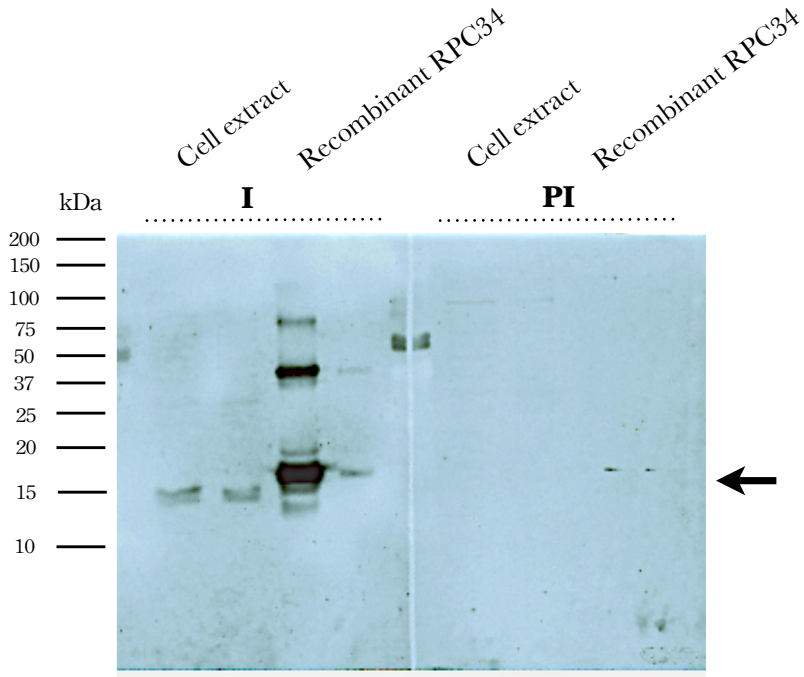


Figure 6.6: *An antibody was raised against recombinant S. acidocaldarius RPC34 in its denatured form in the presence of 4 M urea. Upon incubation of the pre-immune serum control with S. acidocaldarius cell extract and pure, denatured RPC34, no bands corresponding to the size of RPC34 are observed, with the exception of background bands. Upon incubation of the antibody from the final-bleed serum with S. acidocaldarius cell extract and pure, denatured RPC34, similar bands corresponding to the pattern of gel migration of RPC34 of 15 kDa are observed, denoted with an arrow. The higher migration of recombinant, denatured RPC34 is due to the presence of the His-tag. Furthermore, the antibody raised against recombinant, denatured RPC34 also recognises the endogenous RPC34 in the S. acidocaldarius cell extract.*

6.2.4 *In vivo* RPC34 Interaction with the Archaeal RNAP

In order to test for an *in vivo* interaction between RPC34 and the single archaeal RNAP as well as the in vivo relevance of the GST-RPC34 pulldown of the RNAP characterised in terms of Rpo2 subunit, a co-immunoprecipitation (Co-IP) procedure was performed including the newly raised antibodies against RPC34 and Rpo13. First of all, the two antibodies were tested for immunoprecipitation of

their corresponding antigens upon incubation with the *S. acidocaldarius* cell extract, where western blots were performed with the same antibodies to test for antigen-recognition of the immunoprecipitated antigens. As shown in Figure6.7, western blot bands corresponding to RPC34 and Rpo13 of the size of 15 kDa and 20 kDa respectively were observed in the bead fractions when RPC34 and Rpo13 antigens are immunoprecipitated. In contrast bands corresponding to RPC34 and Rpo13 were observed in the supernatant fraction when the immunoprecipitation procedure is carried on with the pre-immune serum. Thus, both RPC34 and Rpo13 antigens are immunoprecipitated by their antibodies bound to protein-A sepharose beads upon incubation with the *S. acidocaldarius* cell extract. In order to test for an in vivo interaction between RPC34 and the RNAP, Rpo13, Rpo2 and RPC34 antigens were immunoprecipitated with their corresponding antibodies, where western blots with different antibodies were subsequently performed in order to test for an interaction between the immunoprecipitated antigen and its corresponding interactant partner. First of all, as observed in Figure6.8, western blot bands corresponding to RPC34 of the size of 15 kDa were observed in the supernatant but not bead fractions when the Rpo13 immunoprecipitation is performed with the pre-immune as well as the immune sera. Secondly, western blot bands corresponding to Rpo13 of the size of 20 kDa were observed in the supernatant but not bead fractions when the RPC34's immunoprecipitation procedure was performed with the pre-immune as well as the immune sera (Figure6.8). Thirdly, western blot bands corresponding to RPC34 and Rpo13 of the size of 15 kDa and 20 kDa respectively were observed in the supernatant but not bead fraction when the Rpo2 antigen was immunoprecipitated by its antibody (Figure6.8). Therefore, no interactions could be obtained between these different protein combinations in these experiments.

To try and reduce the high background signal due to the antibodies and thus increase the sensitivity of the assay, the various co-immunoprecipitation assays were repeated with cross-linked antibodies to the protein-A sepharose beads with subsequent elution of the antigen-interactant partner. As can be observed in

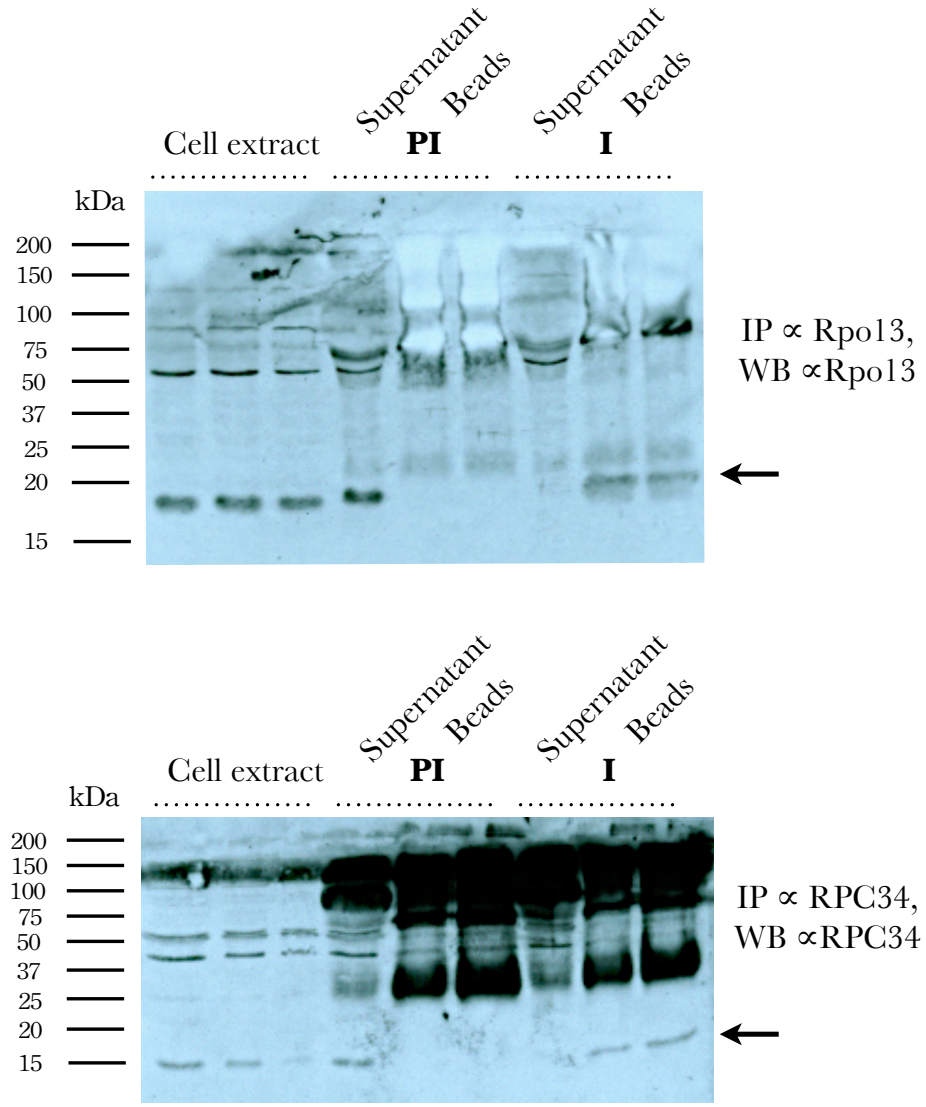


Figure 6.7: *IP assays, where the two antibodies raised against Rpo13 and RPC34 are tested for immunoprecipitation. Western blots are performed with the same antibodies to test for antigen-recognition of the immunoprecipitated antibodies upon incubation with the S. acidocaldarius cell extract. Western blot bands corresponding to Rpo13 and RPC34 of the size of 20 kDa and 15 kDa respectively are observed in the beads fraction when the antigens are immunoprecipitated by their corresponding antibodies in the immune serum (arrows), while bands corresponding to Rpo13 and RPC34 are observed in the supernatant fraction when the immunoprecipitation procedure is performed with the pre-immune serum. The remaining signal corresponds to IgG chains.*

Figure6.9, western blot bands corresponding to RPC34, Rpo13 and Rpo2 of the size of 15 kDa, 20 kDa and 120 kDa were observed in the beads fractions only when the antigen is immunoprecipitated by its corresponding antibody. Furthermore,

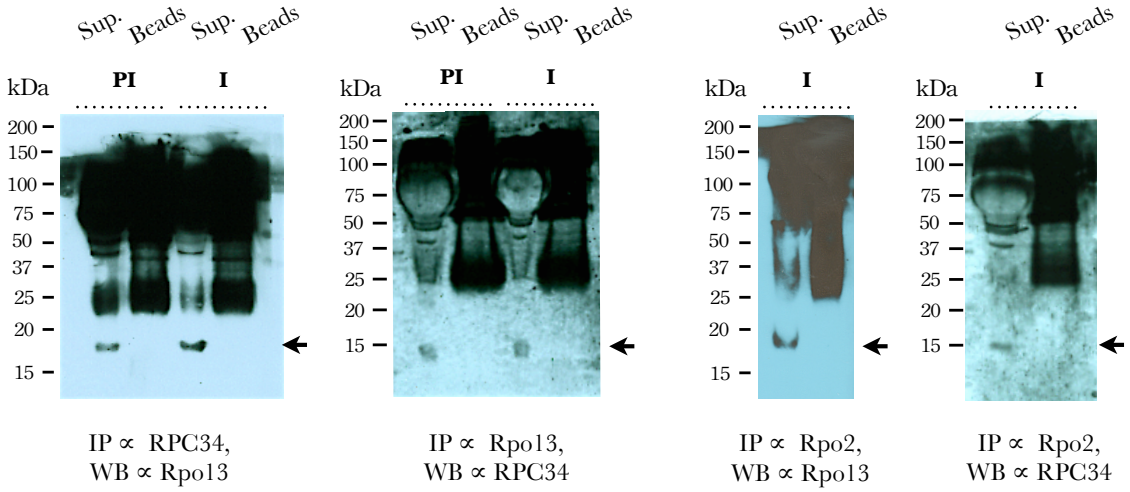


Figure 6.8: *Co-IP assays to test for an interaction between the archaeal RPC34 and the single archaeal RNAP. Upon immunoprecipitation of the antibodies against RPC34 and Rpo13, no co-immunoprecipitation of Rpo13 and RPC34 is observed, as western-blot bands of the size of 20 kDa and 15 kDa respectively are observed only in supernatant samples (arrows). Upon immunoprecipitation of the antibody against Rpo2, no co-immunoprecipitation of either Rpo13 and RPC34 is observed, as western-blot bands of the size of 20 kDa and 15 kDa respectively are observed only in the supernatant samples (arrows). The remaining signal corresponds to IgG chains.*

western blot bands corresponding to RPC34 and Rpo13 of the size of 15 kDa and 20 kDa were observed in the supernatant fractions when Rpo13 and RPC34 antigens are immunoprecipitated, respectively (Figure6.10). In addition, a western blot band corresponding to RPC34 of the size of 15 kDa was observed in the supernatant fraction when the Rpo2 antigen is immunoprecipitated (Figure6.10). Moreover, a western blot band corresponding to Rpo2 of the size of 120 kDa was observed in the beads fraction when the Rpo13 antigen is immunoprecipitated, albeit present at low levels in the supernatant fraction, while it was only present in the supernatant fraction when the immunoprecipitation procedure was performed with the pre-immune serum (Figure6.10). Finally, western blot bands corresponding to Rpo2 of the size of 120 kDa were observed in the supernatant fractions when the immunoprecipitation procedure for the RPC34 antigen was performed with either the pre-immune or immune sera (Figure6.10). Therefore, no co-immunoprecipitation of RPC34 was observed when Rpo13 is immunoprecipitated by its antibody;

no co-immunoprecipitation of Rpo13 and Rpo2 was observed when RPC34 is immunoprecipitated by its antibody; and no co-immunoprecipitation of RPC34 was observed when Rpo2 is immunoprecipitated by its antibody. Finally, partial co-immunoprecipitation of Rpo2 was observed when Rpo13 was immunoprecipitated by its antibody.

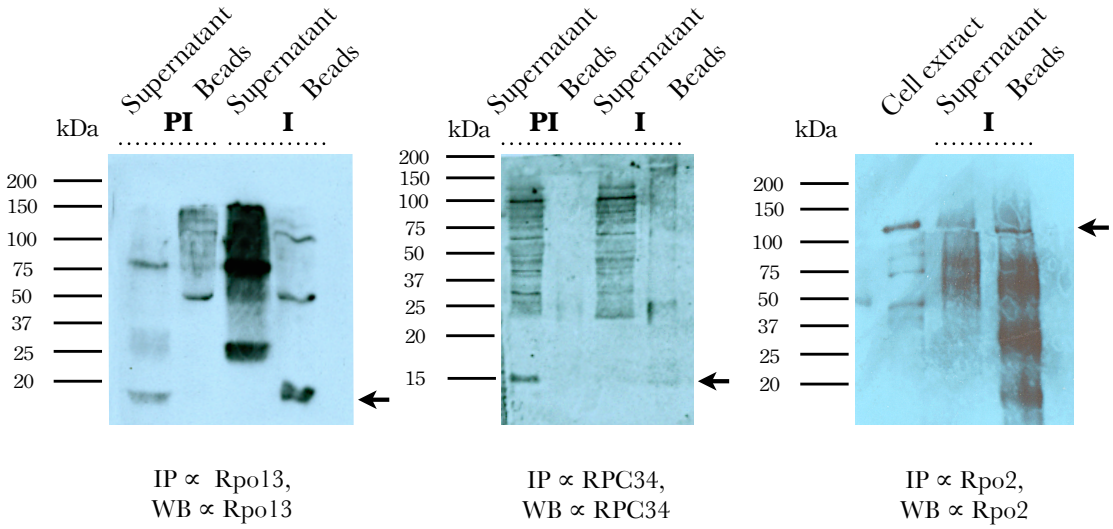


Figure 6.9: *IP assays with cross-linked antibodies to the protein-A sepharose beads, where the two antibodies raised against Rpo13 and RPC34 as well as the antibody against Rpo2 were tested for immunoprecipitation. Western blots were performed with the same antibodies to test for antigen-recognition of the immunoprecipitated antigen upon incubation with the S. acidocaldarius cell extract. Western blot bands corresponding to Rpo13, RPC34 and Rpo2 of the size of 20 kDa, 15 kDa and 120 kDa respectively were observed in the beads fraction when the antigens were immunoprecipitated by their corresponding antibodies in the immune serum (arrows), while bands corresponding to Rpo13 and RPC34 were observed in the supernatant fraction when the immunoprecipitation procedure was performed with the pre-immune serum.*

In order to test for any variations in the level of RPC34 expression throughout the *S. acidocaldarius* cell cycle, a *S. acidocaldarius* 'baby machine' cell synchronisation procedure was performed as previously described [51]. This procedure consists of synchronising cells at the start of the G1 phase, where in the 'baby machine', an asynchronously dividing culture of cells bound to the surface of a membrane is continually perfused with medium. As cells divide, one daughter cell remains bound to the membrane and the other falls into the liquid medium. The liquid eluate therefore contains early-G1 synchronised cells that have a very high degree

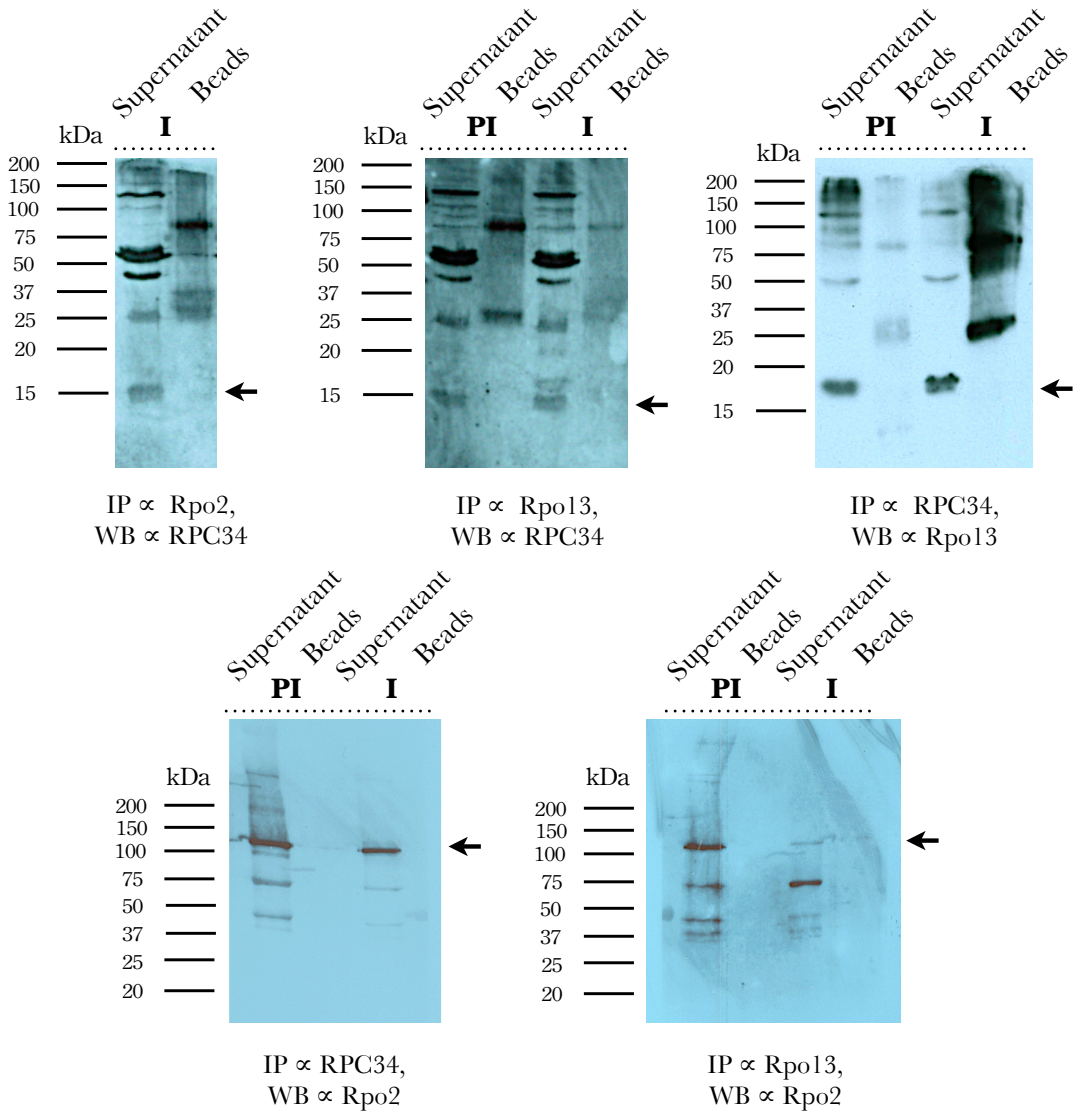


Figure 6.10: *Co-IP assays with cross-linked antibodies to the protein-A sepharose beads in order to test for an interaction between the archaeal RPC34 and the single archaeal RNAP. Upon immunoprecipitation of the antibodies against RPC34 and Rpo13, no co-immunoprecipitation of Rpo13 and RPC34 is observed, as western-blot bands of the size of 20 kDa and 15 kDa respectively are observed only in supernatant samples (arrows). Upon immunoprecipitation of the antibody against Rpo2, no co-immunoprecipitation of RPC34 is observed, as western-blot bands of the size of 15 kDa are observed only in the supernatant samples (arrows). Finally, while upon immunoprecipitation of the antibody against RPC34, no co-immunoprecipitation of Rpo2 is observed, as western-blot bands of the size of 120 kDa are observed only in the supernatant samples (arrows), upon immunoprecipitation of the antibody against Rpo13, partial co-immunoprecipitation of Rpo2 is observed, as western-blot bands of the size of 120 kDa are observed in the beads fraction in addition to the supernatant samples (arrows).*

of cell synchrony. As can be seen in Figure6.11, upon fluorescence-activated cell sorting (FACS) analysis of 'baby machine' cell synchronised samples collected after 2 hours, 3 hours, 4 hours and 5 hours of 'baby machine' cell synchronisation, the best *S. acidocaldarius* cell synchronisation at G1 was obtained after 3 hours, as previously reported [51]. Upon growth and sample collection at various time points of synchronised *S. acidocaldarius* cell culture collected after 3 hours of 'baby machine' procedure, RPC34 (15 kDa) as well as TBP (19 kDa) protein expression levels in the various collected samples were analysed by western blots performed with the corresponding antibodies, where TBP protein expression levels are known to be constant throughout the cell cycle [150]. As can be seen in (Figure6.13), while FACS analysis showed full conversion of the *S. acidocaldarius* cell culture in the G1 phase into the G2 phase (Figure6.12), no difference in the RPC34 and TBP protein levels were found to occur throughout the cell cycle, with the exception of steadily increasing expression levels of both proteins corresponding to growing cells.

6.2.5 RPC34's Interaction with the Archaeal TFB-2

In order to test for an interaction between RPC34 and the paralog of TFB referred to as TFB-2, a yeast two-hybrid screen was performed. In this screen, RPC34 is encoded in fusion with either the Gal4 activation domain or the Gal4 binding domain, while TFB as well as TFB-2 are encoded in fusion with either the Gal4 activation domain or the Gal4 binding domain. As can be seen in Figure6.14, all yeast strains grow in the permissive *-trp* and *-leu* conditions, which selected for both coding plasmids. Furthermore, no selective yeast growth in the *-his* condition was observed upon yeast co-transformation with any of the plasmids, with the exception of yeast growth obtained upon the previously reported interaction between Gins15 and RecJ components of the DNA replication machinery (Figure6.4) [118]. No yeast growth was observed upon yeast co-transformation with empty plasmids.

In order to test for the expression of the various constructs in the co-transformed yeast strains, yeast protein extracts were prepared and subjected to western

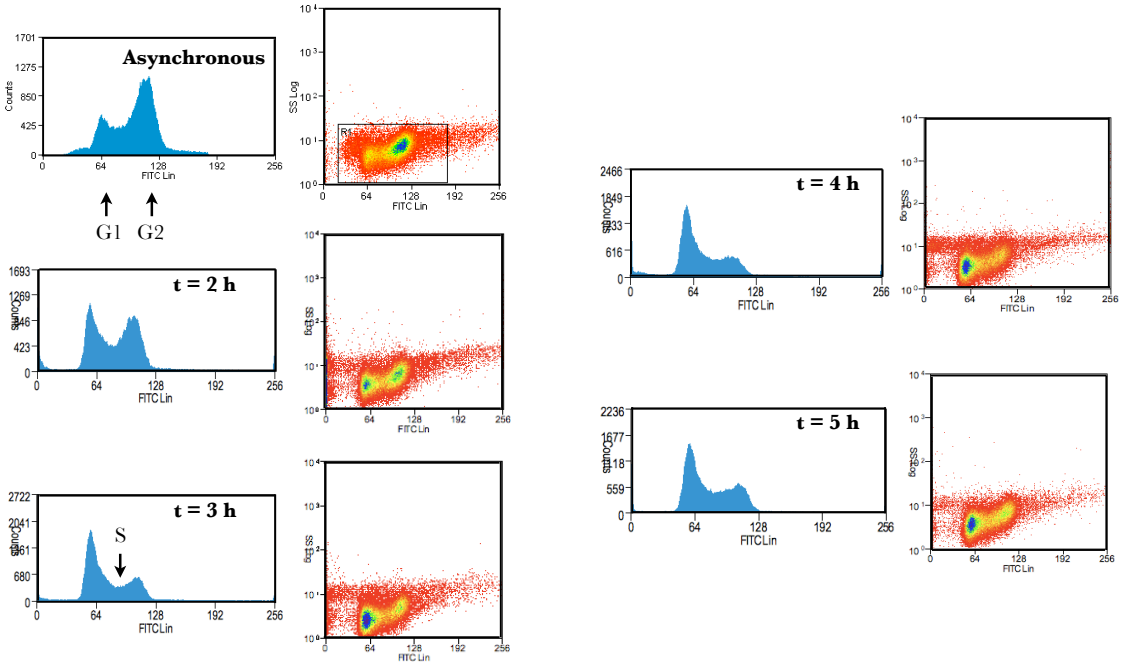


Figure 6.11: *A 'baby machine' cell synchronisation procedure run, where enriched G1 S. acidocaldarius cell samples are collected prior to sample growth, and after 2 h, 3 h, 4 h and 5 h of synchronisation procedure. The samples are subsequently analysed on a FACS machine for DNA content, where the first peak represents the percentage of the cell population in G1 phase, while the second peak represents the percentage of the cell population in G2 phase (arrows). The lower the convex shape, the lower the cell population in S phase (arrow). The best synchronisation of the S. acidocaldarius cell population is achieved after 3 h, with the highest G1 peak and the lowest convex shape.*

blot analysis with antibodies against the HA-epitope as well as the Myc-epitope expressed in fusion with all of the proteins in conjunction with the antibody against RPC34. As can be seen in Figure6.15, the antibody against the HA-epitope recognised a non-specific yeast protein also present in the plasmid-free AH109 control, showing unspecific antibody binding. Furthermore, the antibody against the Myc-epitope recognised bands corresponding to the size of the Myc peptide fragment of around 20 kDa for the yeast strains having empty pGBK-T7 plasmid, while bands corresponding to the size of the Myc-labelled RPC34 of around 35 kDa for the yeast strains having the RPC34-pGBK-T7 plasmid (Figure6.15). Moreover, bands corresponding to either the size of the Myc-labelled TFB of around 55 kDa for the yeast strains having the TFB-pGBK-T7 plasmid or bands corresponding to

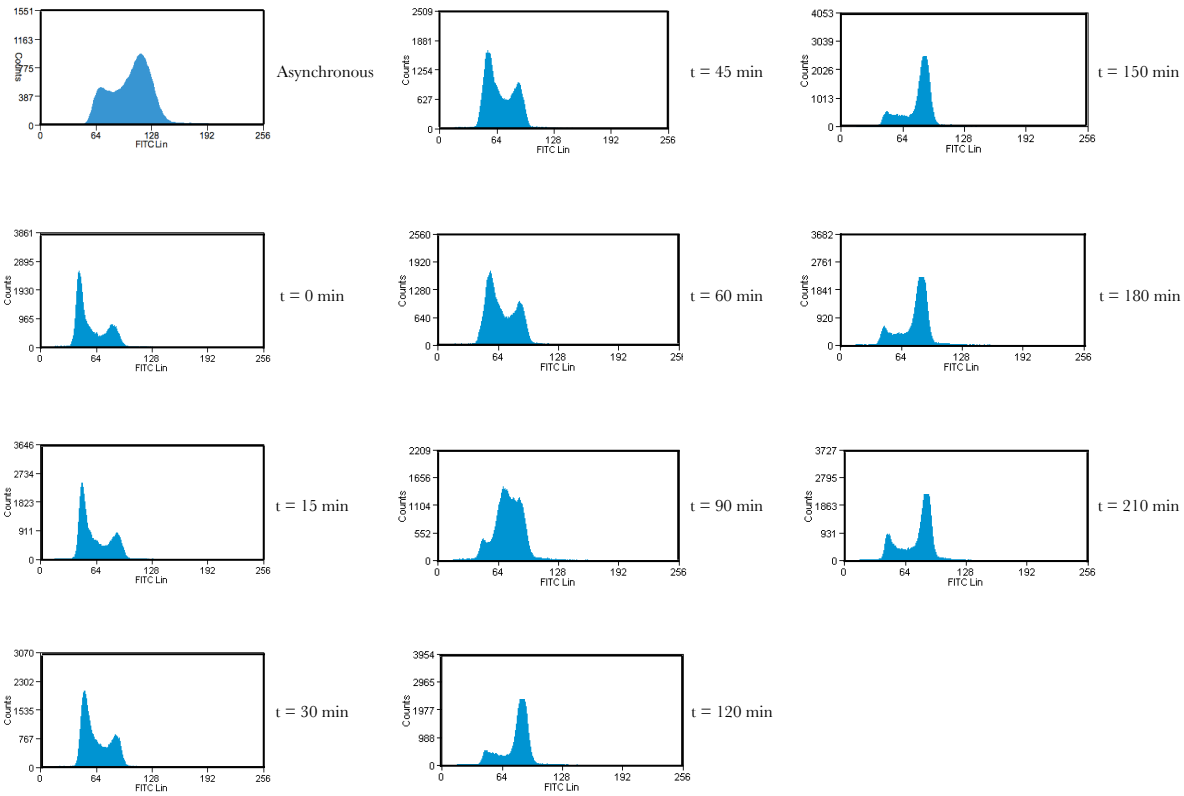
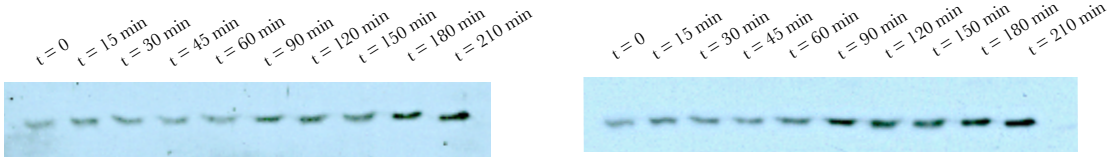


Figure 6.12: *A synchronised S. acidocaldarius cell culture sample collected after 3 h of 'baby machine' cell synchronisation procedure was grown for 210 mins, where samples were collected at the indicated time points. The samples were subsequently analysed on a FACS machine for their DNA profiles. The cell cycle profiles of the various S. acidocaldarius cell samples go from the one with the highest G1 peak at t = 0 min to the one with the highest G2 peak at t = 150 min. Thus, the analysis covers the entire S. acidocaldarius cell cycle profile.*



RPC34 levels throughout the cell cycle

TBP levels throughout the cell cycle

Figure 6.13: *TBP and RPC34 protein expression levels were analysed with their corresponding antibodies in the various samples collected at the various time points after growth of a synchronised S. acidocaldarius cell culture sample. Western blot bands of the size of 19 kDa and 15 kDa are obtained for TBP and RPC34 respectively. In both TBP and RPC34 cases, a steady increase in the protein levels is observed for the collected samples, corresponding to growing cells. Thus, no fluctuations in the RPC34's protein levels occur throughout the S. acidocaldarius cell cycle.*

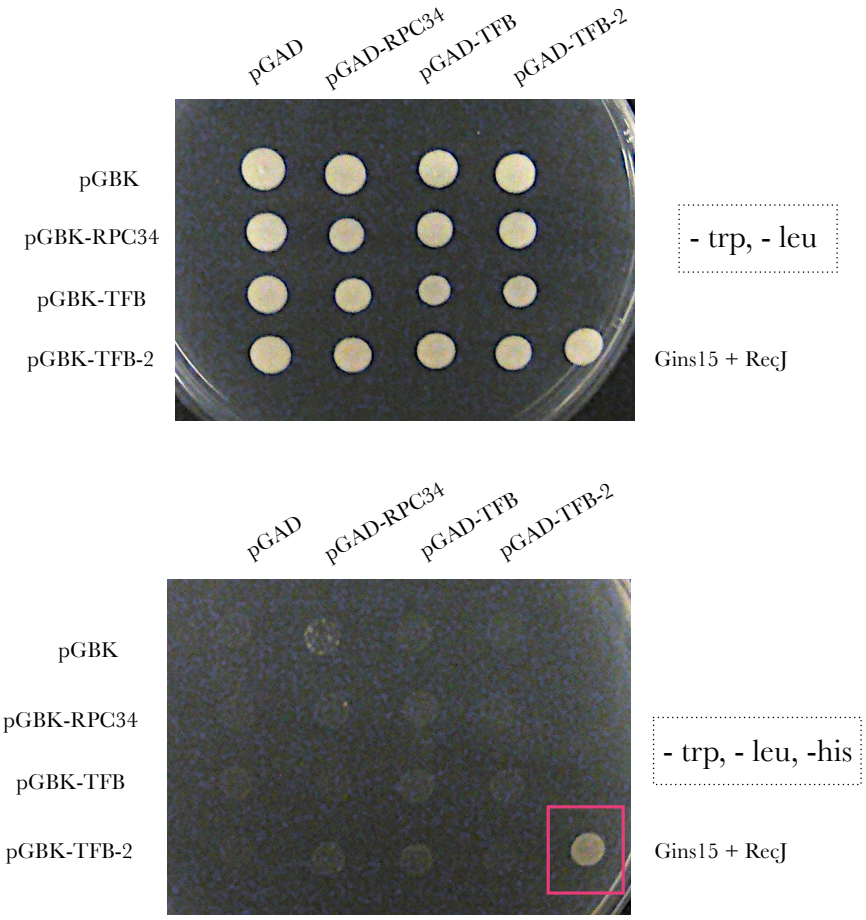


Figure 6.14: *Yeast two-hybrid screen, where an interaction between RPC34 and TFB-2 is tested. Yeast strains co-transformed with plasmids coding for either RPC34 or empty vector in conjunction with TFB and TFB-2 grow the in the permissive -trp and -leu condition, which selects for pGBK and pGAD plasmids. However, upon growth in the selective -trp, -leu and -his condition, only the yeast strains co-transformed with plasmids coding for Gins15 and RecJ (pink box), used as a positive control, are observed to grow. No selective growth occurs upon yeast co-transformation with either pGBK-RPC34 and pGAD-TFB-2 plasmids, or pGAD-RPC34 and pGBK-TFB-2 plasmids.*

the size of the Myc-labelled TFB-2 of around 53 kDa for the yeast strains having the TFB-2-pGBK-T7 plasmid were observed (Figure6.15). Finally, bands corresponding to the size of the RPC34 protein of around 15 kDa for the yeast strains having either the RPC34-pGBK-T7 plasmid or RPC34-pGAD-T7 plasmid were also observed.

In addition, RPC34's interaction with TFB-2 was tested by means of co-expression. Indeed, it is found that not only RPC34 is insoluble when Ni^{2+}-NTA affinity purified with standard conditions, but also TFB-2 is insoluble when

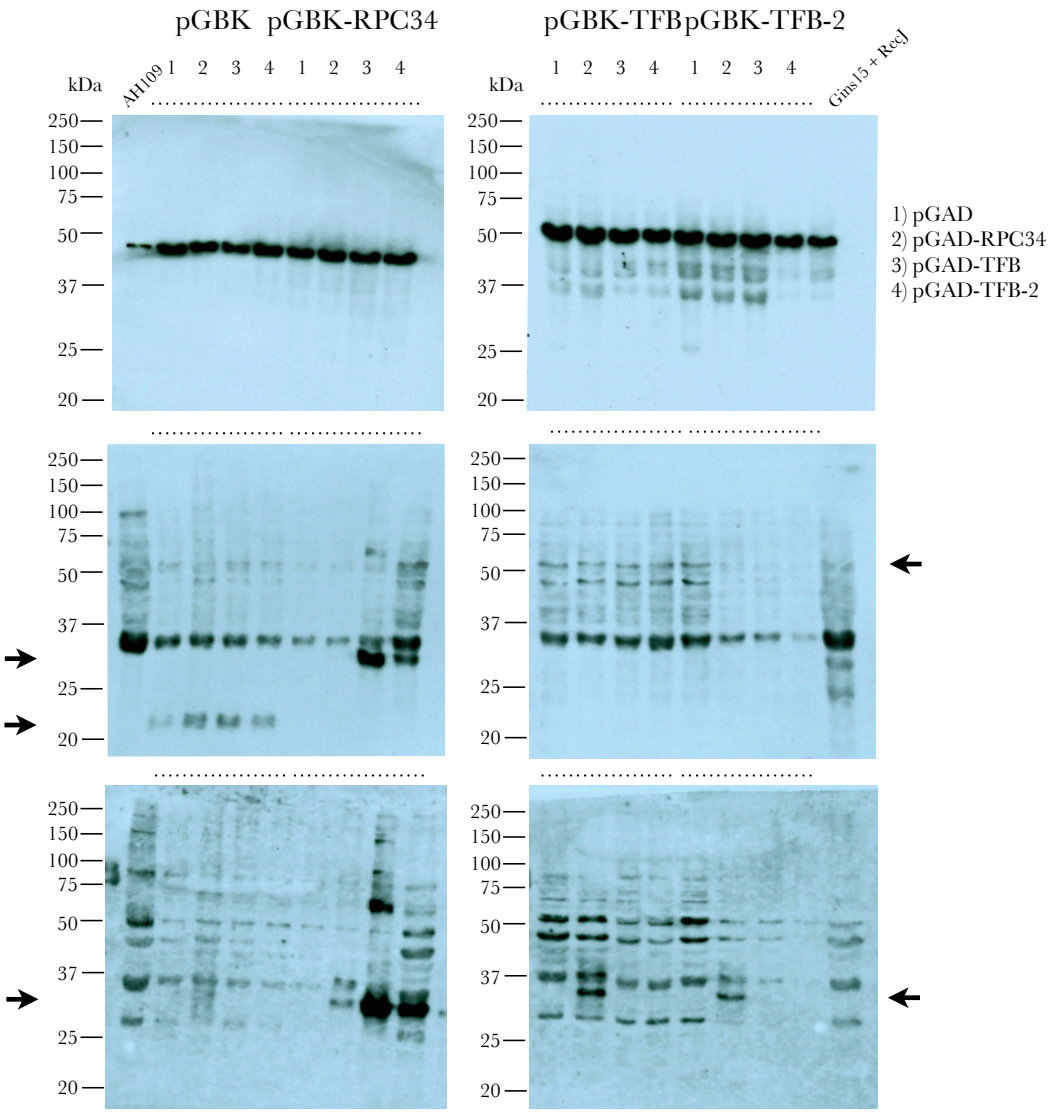


Figure 6.15: *Analysis of the expression of the various constructs in the co-transformed yeast strains, where western blots were performed on yeast protein extracts. The antibody against the HA-epitope found in pGAD plasmids recognises a non-specific yeast protein also present in the plasmid-free AH109 control. The antibody against the Myc-epitope found in pGBK plasmids recognises bands corresponding to the size of the Myc peptide fragment of around 20 kDa for the yeast strains having empty pGBK-T7 plasmid (arrow). Furthermore, it recognises bands corresponding to the size of the Myc-labelled RPC34 of around 35 kDa for the yeast strains having the RPC34-pGBK-T7 plasmid (arrow). Moreover, it also recognises bands corresponding to either the size of the Myc-labelled TFB of around 55 kDa for the yeast strains having the TFB-pGBK-T7 plasmid or bands corresponding to the size of the Myc-labelled TFB-2 of around 53 kDa for the yeast strains having the TFB-2-pGBK-T7 (arrow). Finally, the antibody against RPC34 recognises bands corresponding to the size of the RPC34 protein of around 15 kDa for the yeast strains having either the RPC34-pGBK-T7 plasmid or RPC34-pGAD-T7 plasmid.*

expressed in its entirety or as a core domain (Figure6.16). It is found that 4 M of guanidinium hydrochloride is required for Ni^{2+}-NTA affinity purification of both the entire and core TFB-2 proteins under denaturing conditions (Figure6.16), where guanidinium hydrochloride is used as opposed to urea since it has ionic properties and it eliminates any destabilising electrostatic interactions from the charge of TFB-2 patches. In order to test whether RPC34 interacts with TFB-2, and whether this interaction leads to the formation of a soluble RPC34-TFB-2 complex, both RPC34 and TFB-2 were cloned into compatible co-expression plasmids. If solubilised S-tagged RPC34 is affinity purified with S-beads, soluble TFB-2 should also co-elute, since it would be part of the soluble RPC34-TFB-2 complex. As can be seen in Figure6.16, no bands corresponding to the expressed RPC34 and TFB-2 of the size of 15 and 33 kDa were observed as soluble components in either the supernatant fraction after cell lysis, or in the elution fraction after elution from the S-beads.

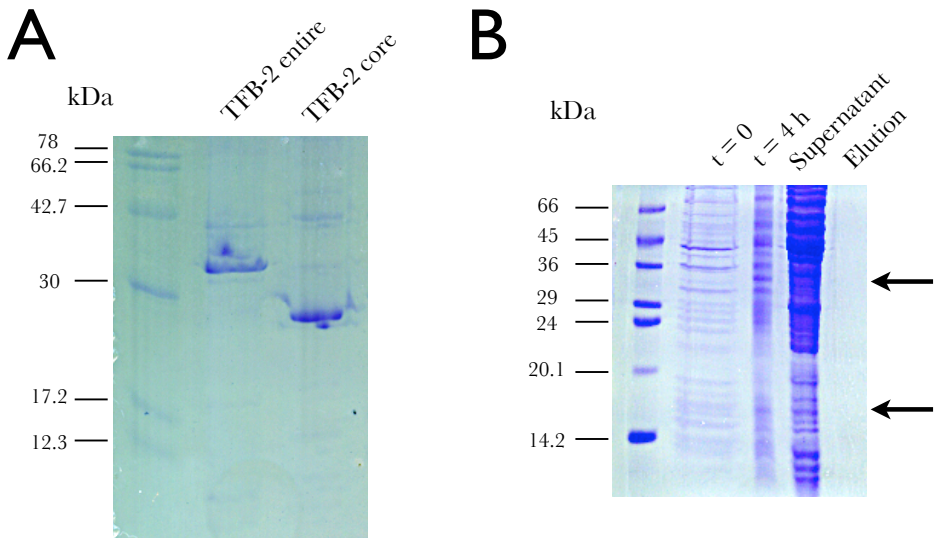


Figure 6.16: *RPC34's interaction with TFB-2 tested by means of co-expression.* ***(A)*** *TFB-2 expressed either entirely or as a core domain requires 4 M of guanidinium hydrochloride for its solubilisation, since it is expressed in the E. coli inclusion bodies.* ***(B)*** *Upon co-expression of plasmids coding for either RPC34 or TFB-2, expression products are observed of the size of 15 kDa and 33 kDa respectively (arrows). However, upon analysis of the supernatant fraction after cell disruption and analysis of the elution fraction after S-tagged RPC34 affinity purification upon passing through S-beads, no observable bands of the size of 15 kDa and 33 kDa are observed. Thus, no soluble RPC34-TFB-2 complex is formed upon co-expression of insoluble RPC34 and insoluble TFB-2 proteins together.*

6.3 Discussion

The archaeal RPC34 homologue is proposed to interact with the single archaeal RNAP, thus defining a new RNAP, which would act as a precursor of the eukaryotic RNAP-III, thereby modulating its specificity for the transcription of a subset of genes, such as non-coding rRNA and tRNA genes. Furthermore, a TFB paralog referred to as TFB-2 could be also involved in directing the transcription of this subset of non-coding rRNA and tRNA genes. TFB-2 could interact with RPC34 similarly to the RNAP-III system, where RPC34 interacts with TFIIIB70, and the TFB-2-RPC34 complex could be involved in the recruitment of the archaeal RNAP to structural promoters. In addition to RNAP recruitment, RPC34 could also have a regulatory role via direct interaction with the DNA, since different transcription levels from the structural 16S/23S rRNA promoter are obtained by the reconstituted system and crude extract [138].

In order to test for an interaction between RPC34 and the single RNAP, a GST-pull down assay was performed, where a band corresponding to the Rpo2 subunit of the size of 120 kDa was observed only when the RPC34-GST fusion but not the GST protein alone was incubated with the *S. acidocaldarius* cell extract and not the buffer control. This suggests that RPC34 is able to interact with the RNAP, and to 'pull it down' upon the binding of RPC34-GST to the glutathione-sepharose beads. To test whether RPC34 interacts with any specific RNAP's subunits, a yeast-two-hybrid screen was set up where selective yeast growth occurs upon specific interaction of RPC34 with the Rpo8 subunit of RNAP. This interaction was found to be weak, as yeast growth disappears upon introduction of the 3AT inhibitor of the product of the *his3* gene.

Furthermore, in order to find the *in vivo* relevance of the GST-pull down of the RNAP by RPC34-GST, co-immunoprecipitation assays were set up, where it is found that Rpo13 and RPC34 proteins are immunoprecipitated by their corresponding antibodies. Therefore, the raised antibodies against RPC34 and Rpo13 are capable

of interacting with their antigens when immunoprecipitated, where the antibody raised against denatured RPC34 also recognises properly folded, endogenous RPC34 from the *S. acidocaldarius* cell extract. On the other hand, when the interaction between RNAP and the interactant RPC34 partner is tested, it is found that RPC34 does not co-immunoprecipitate with either Rpo13 or Rpo2, and Rpo13 does not co-immunoprecipitate with RPC34. In order to increase the sensitivity of the co-immunoprecipitation assays and to allow for the detection of western blot bands corresponding to Rpo2, antibodies were cross-linked to the protein-A sepharose beads. It was found that Rpo13, RPC34 and Rpo2 proteins were immunoprecipitated by their corresponding antibodies. Furthermore, it was found that RPC34 does not co-immunoprecipitate with either Rpo13 or Rpo2, and that neither Rpo13 nor Rpo2 co-immunoprecipitate with RPC34. All together, these results argue against an interaction between RPC34 and RNAP from an *in vivo* perspective. It could be that the interaction between the RNAP and RPC34 is weak in its nature and therefore prone to dissociation. Alternatively, only a very minor fraction of RNAP from the general RNAP pool is bound to RPC34 *in vivo*, and therefore the interaction is hard to detect. Finally, the interaction could be of a transient nature such as cell-cycle regulated, since TFB-2 expression levels were found to fluctuate according to the cell cycle upon microarray mRNA expression analysis (Rachel Samson, personal communication). In order to test for cell cycle regulation of RPC34's expression levels, a baby-machine approach was undertaken. It is found that, upon sample analysis by means of a western blot with the antibody against RPC34 at various time points of a synchronised cell culture, no fluctuations in the RPC34 levels throughout the cell cycle occur, as with TBP levels. Only steadily increasing RPC34 levels were observed with time, explained by the increasing cell number. Thus, no cell cycle regulation of RPC34's protein levels occur. In addition, Rpo13 did not to co-immunoprecipitate with Rpo2. The western blot band corresponding to Rpo13 of the size of 20 kDa observed in the supernatant fraction could correspond to the free Rpo13 pool that is not associated

with the RNAP, as suggested by the gel filtration experiment discussed in Chapter 1, in the section 'Rpo13's *in vivo* Association with the Archaeal RNAP within the *S. acidocaldarius* Cell Extract'. Furthermore, RNAP's binding to the antibody against Rpo2 most likely occludes the Rpo13 epitope for the antibody against Rpo13, explaining why no band corresponding to Rpo13 is observed in the in the beads fraction when Rpo2 is immunoprecipitated. On the other hand, Rpo2 was found to co-immunoprecipitate with Rpo13, suggesting that RNAP's binding to the antibody against Rpo13 partially occludes the Rpo2 epitope for the antibody against Rpo2, since a fraction of Rpo2 protein is not only found in the beads sample but also the supernatant sample. To conclude, overall these results argue against an *in vivo* interaction between RPC34 and the single archaeal RNAP. It is possible that the observed *in vitro* association between RPC34-GST and the RNAP was favoured by the large amount of added RPC34-GST protein, which is in vast excess with respect to the cellular and endogenous RNAP from the *S. acidocaldarius* cell extract, and so does not reflect the *in vivo* situation. Also, the assay was performed at 30 °C, which is a suboptimal temperature and may promote an unspecific interaction. Furthermore, selective yeast growth upon interaction between RPC34 and Rpo8 might be due to unspecific binding of components such as the Gal4 binding domain fused with RPC34. Indeed, selective growth was observed upon interaction between the Gal4 binding domain on its own and the Rpo3, Rpo4, Rpo6, Rpo7 and Rpo12 subunits on their own, suggesting that the Gal4 binding domain might account for unspecific binding. Since no experiments were performed to test for the expression of fusion proteins in yeast, it is not possible to definitely assess the existence of true interactions.

In order to test for an interaction between RPC34 and TFB-2, a yeast-two-hybrid screen as well as co-expression of recombinant plasmids was performed. Indeed, RPC34's interaction with TFIIIB70 was shown by *in vitro*, where a GST pull-down experiment showed the interaction ([91], [169]), and of all the RNAP-III subunits investigated, only RPC34 was found to interact with TFIIIB70 in the two-hybrid

system [173]. Upon yeast-two-hybrid screen of RPC34 versus TFB and TFB-2, no selective yeast growth is observed, showing that no interaction between RPC34 and TFB-2 occurs. To discard the hypothesis of a problem in the expression of the constructs in yeast, yeast protein extracts from all the various strains were tested for expression of the constructs. It is found that all the yeast strains have the appropriate expression of constructs, with the antibody against RPC34 being more sensitive in protein detection than the Myc-antibody. Therefore, the absence of observed interaction between RPC34 and TFB-2 investigated by yeast-two-hybrid suggests that either the interaction does not exist, is too weak and prone to dissociation, is of a transient nature, or additional components are needed to bridge such interaction. To observe the existence of such RPC34-TFB-2 interaction, co-expression plasmids coding for either RPC34 or TFB-2 respectively, which are insoluble when individually expressed, were tested for expression of a soluble RPC34-TFB-2 complex when *E. coli* cells were co-transformed with both plasmids. It was found that such co-expression does not lead to the formation of a soluble RPC34-TFB-2 complex, as no soluble RPC34 is purified by affinity chromatography with S-beads, with no co-elution of TFB-2 with RPC34. Therefore, solubilisation of RPC34 with TFB-2 does not occur upon co-expression of the two components. It is possible that additional components are needed to lead to the formation of a soluble complex. All together, these results argue against an interaction between RPC34 and TFB-2.

In summary, while it has been postulated that the single archaeal RNAP interacts with RPC34 in conjunction with TFB-2 to direct the transcription from a subset of genes such as non-coding rRNA and tRNA genes, the results obtained argue against the existence of an interaction between RPC34 and RNAP, and an interaction between RPC34 and TFB-2. Nonetheless, it is possible that the employed methods are not sensitive enough for the detection of these interactions, since protein-protein interactions can be weak, temporary, dependent on protein amounts or subject to regulation such as cell cycle regulation. Furthermore, since RPC34 has been

proposed to directly contact DNA by means of cross-linking analysis and to have an effect in open complex formation [29], [11], RPC34 might act at the DNA level rather than at a protein level through RNAP interaction. A complementary analysis consists of analysing the distribution of RPC34 at the level of the *S. acidocaldarius* genome with respect to the RNAP distribution. This analysis would allow determination of whether RPC34 is localised at a subset of genes such as non-coding rRNA and tRNA genes, and therefore could infer that RPC34 has a regulatory role in their expression through DNA binding.

Chapter 7

Analysis of Rpo13 Genomic Distribution

7.1 Introduction

In this study, the Rpo13 subunit specific to the archaeal RNAP was found to interact with the DNA, where its C-terminal domain is proposed to mediate DNA binding [179]. While the *S. shibatae* RNAP crystal structure shows that Rpo13 is a *bona fide* RNAP subunit, Rpo13's binding to the DNA and its putative role in transcription, where Rpo13 might help to unwind the DNA through DNA contacts with its N- and C-termini, has led to the proposal that Rpo13 might be a functional subunit that exerts an equivalent role to the eukaryotic initiation factors absent in Archaea. Indeed, transcription of a bubble DNA oligo incubated with purified RNAP occurs, where in the RNAP-DNA crystal structure this bubble DNA oligo is shown to have contacts with the Rpo13 N- and C-termini after the bubble region ([180]). The idea that Rpo13 might behave more as a transcription factor than a subunit with only a structural role is further supported by the fact that the $\alpha 2$ helix of Rpo13 is characterised by polar residues, showing some similarity to the charged region of Tfg1-like factors of TFIIF [180]. Furthermore, in solution Rpo13 forms mainly dimers [179], a property also shown by TFIIF, which has a specific

dimerisation domain preceding the charged region [61], [39]. These observations have led to the analysis of gel filtration elution volumes of endogenous Rpo13 from the *S. acidocaldarius* cell extract with respect to that of recombinant Rpo13. It was found that endogenous Rpo13 from the *S. acidocaldarius* cell extract, despite predominantly co-eluting with the Rpo2 subunit at an elution volume corresponding to the one of purified RNAP on its own, also elutes at a later elution volume corresponding to a simple Rpo13 molecule. These results suggest that, while the endogenous Rpo13 from the general cellular pool is mainly associated with the archaeal RNAP, a fraction of unbound Rpo13 also exists. Thus, it is possible that Rpo13 dissociates from the RNAP after performing a transcription-like role, leading to the existence of the free Rpo13 pool. Rpo13 might therefore be defined as a transcription factor which is only transiently associated with the RNAP as opposed to a *bona fide* RNAP subunit which is constitutively and permanently associated with the RNAP.

In order to define Rpo13 in terms of either subunit or transcription factor, a more in depth *in vivo* analysis of its distribution within the *S. acidocaldarius* genome with respect to that of the general RNAP pool in its elongation form is required. Indeed, all Rpo13 might be either associated *in vivo* with the general RNAP pool while transcribing genes, or a subset of RNAP may lack Rpo13, leading to a general free Rpo13 pool. Therefore, Rpo13 could be associated with the RNAP only when the RNAP is found at promoters during transcription initiation, where Rpo13 could perform some of the roles equivalent to eukaryotic transcription factors absent in Archaea. Upon promoter clearance, Rpo13 could dissociate from the RNAP, leading to an elongating RNAP lacking Rpo13.

In order to allow for an *in vivo* analysis of the distribution of Rpo13 within the *S. acidocaldarius* genome with respect to that of the general RNAP pool at promoters of various genes as well as in its elongation form, chromatin immunoprecipitation with the antibodies raised against Rpo2 and Rpo13 followed by sequencing of the immunoprecipitated DNA fragments (ChIP-seq) is the technique of choice. Indeed,

this technique measures the distribution of signal around the genes. This reduces the fact that coupled transcription and translation in Archaea ([58]) would also lead to cross-linked DNA-RNA-protein products, where the protein would be found at the position of its coding gene, hampering the analysis of the various signals obtained from independent and authentic association events of protein to its various binding sites on the DNA. The distribution of the ChIP-seq signal around individual genomic loci would allow an assessment of the protein's broad *in vivo* occupancy at the various loci. If its occupation is RNA-mediated, the distribution of the signal would be very narrow, as it would be determined only by DNA-RNA base pairing over a short DNA region. In conjunction with ChIP-seq performed to analyse Rpo13's distribution versus RNAP, sequencing of the various RNA products (RNA-seq) would lead to a more global analysis and would correlate the abundance of transcript levels with RNAP localisation. All together, these analyses should yield novel insights into the dynamics of RNAP localisation and transcription activity.

7.2 Results

In order to analyse the *in vivo* genomic distribution of Rpo13 with respect to that of the bulk RNAP, the antibody raised against Rpo13 was used in the ChIP-seq approach. As can be seen in Figure7.1, protein A-Sepharose beads bind to the antibody raised against Rpo13 in the buffer used for the ChIP-seq procedure consisting in TBS buffer with 0.5% Tween-20 and 0.5% Triton X-100. For these ChIP sequencing experiments in this chapter, the crosslinked cell extract was prepared in conjunction with Dr. Priscilla Braglia.

7.2.1 Rpo13's *In Vivo* Genomic Localisation versus the One of the RNAP

In order to analyse the *in vivo* genomic distribution of Rpo13 with respect to that of the bulk RNAP, a ChIP-seq approach was undertaken with the antibodies raised

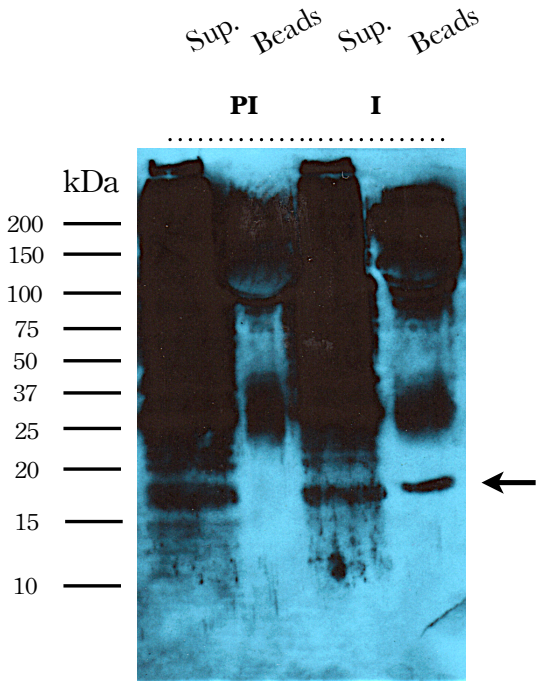


Figure 7.1: *IP assays, where protein A-Sepharose beads bind to the antibody raised against Rpo13 in the buffer used for the ChIP-seq procedure. The buffer consists in TBS buffer with 0.5% Tween-20 and 0.5% Triton X-100, the mild detergents potentially causing protein denaturation or hampering the assay by disrupting the antibody's interaction with either protein A-Sepharose beads or its corresponding antigen.*

against Rpo13 as well as Rpo2. The analysis of the genomic distribution of Rpo13 will reveal whether Rpo13 always co-localises with Rpo2, or whether it co-localises with Rpo2 predominantly at certain genomic regions, such as promoters. The various DNA libraries, generated by Source BioScience, were obtained upon adding Illumina Indexing adapters on both ends and subsequent 17 cycles of PCR to amplify the amount of prepared material, where the average size of the libraries is 300 bp. Samples were then sequenced using 55 bp single-end runs with a 7 bp index read. 2,011,670, 2,217,640 and 4,474,737 reads were respectively obtained for the Rpo13, Rpo2 and input samples, and of these 106,153 reads, 3,512,103 reads and 129,964 reads were mapped onto the *S. acidocaldarius* genome.

Probe values were assigned to the number of reads mapped to each gene in the *S. acidocaldarius* genome, where contigs closer than 30 bp are merged, the minimum contig size is 250 bp corresponding to the average *S. acidocaldarius* open reading

frame, and a value of 0.8 is assigned for a decent signal to noise coverage defined as depth cutoff. As can be seen in Figure7.2, a total significant number of 1080 probes were obtained and assigned to the number of reads mapped to each gene in the *S. acidocaldarius* genome for both Rpo2 and Rpo13. All probe values were standardised to the input sample. As shown in Figure7.3, the validity of the probes were assessed by an independent real-time Q-PCR experiment with primers designed against the ribosomal 5S rRNA coding region, the thermosome coding region and a region containing no probe values defined as a negative region. It was found that the regions corresponding to the ribosomal 5S rRNA coding region and the thermosome coding region are amplified by the designed primers in both the immunoprecipitated protein-DNA complexes by the antibodies against Rpo2 and Rpo13, while little amplification occurred for the negative region. The Q-PCR data therefore support the idea that the probe value assignment is an accurate representation of the *in vivo* genomic occupancy of Rpo2 and Rpo13. Furthermore, in order to correlate the Rpo2 distribution with Rpo13, a correlation plot was obtained with the two data sets (Figure7.2). As can be seen by the correlation R factor value of 0.9, there is a very good overall and global correlation between Rpo2 and Rpo13 subunits. Therefore, Rpo13 is not sub-located to specific coding regions independently from Rpo2, such as at the 5' ends of genes, but instead co-localises with the general RNAP pool in its elongation form and overall transcription cycle. Moreover, upon analysis of the highly transcribed 50S ribosomal protein genes cluster (Figure7.4), there was also a very good localised correlation between Rpo2 and Rpo13 subunits, where a running window probe generator function with a window size and step size of both 20 bp was applied to the reads. Therefore, in summary, there is a very good correlation between Rpo2 and Rpo13 subunits at both at a global and local level.

Finally, a running window probe generator function with a window size and step size of both 20 bp was applied to the Rpo2 and Rpo13 reads in order to analyse the RNAP's distribution across individual *S. acidocaldarius* genes. As shown in Figure7.5, the RNAP occupies mainly the middle of the genes in most cases, while

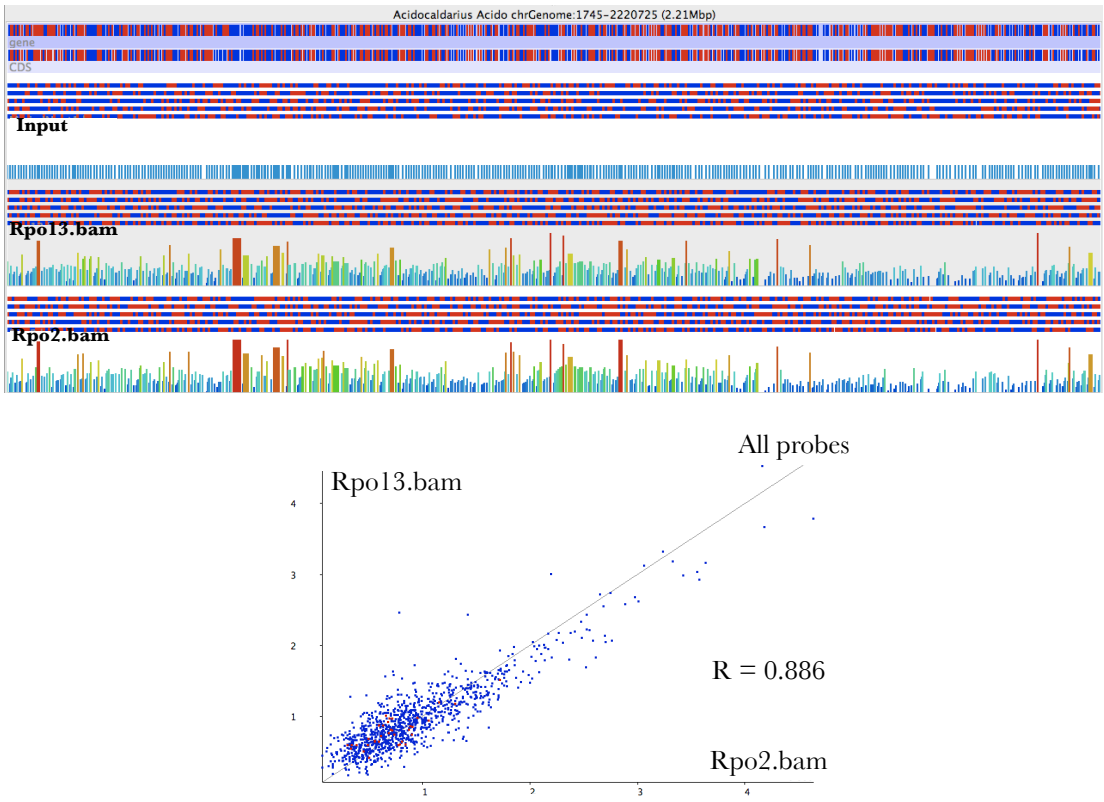


Figure 7.2: *ChIP-seq experiments performed with the antibodies against the RNAP's Rpo2 and Rpo13 subunits. All the values of the obtained 1080 probes are standardised to the input sample. There is a very good overall, global correlation between Rpo2 and Rpo13 subunits, as indicated by the correlation R factor value of 0.9 obtained upon plotting probes assigned to Rpo13's genomic positions versus probes assigned to Rpo2's genomic positions. Thus, Rpo13 co-localises with the general RNAP pool at both transcription initiation and elongation.*

it is found either at the 5' or 3' end of genes only in a minority of cases.

7.2.2 Transcript Abundance versus RNAP Localisation

In order to analyse the *in vivo* abundance and levels of the various RNA transcripts with respect to that of the RNAP's distribution at the various genes, an RNA-sequencing (RNA-seq) approach was undertaken in parallel with ChIP-seq with the antibodies raised against Rpo13 and Rpo2. As shown in Figure7.6, a RNA sample from the *S. acidocaldarius* cell extract was prepared. The sample was found to contain the predominant 23S, 16S and 5S rRNA species (Figure7.6). The RNA library, generated by Source BioScience, was prepared similarly to the DNA library,

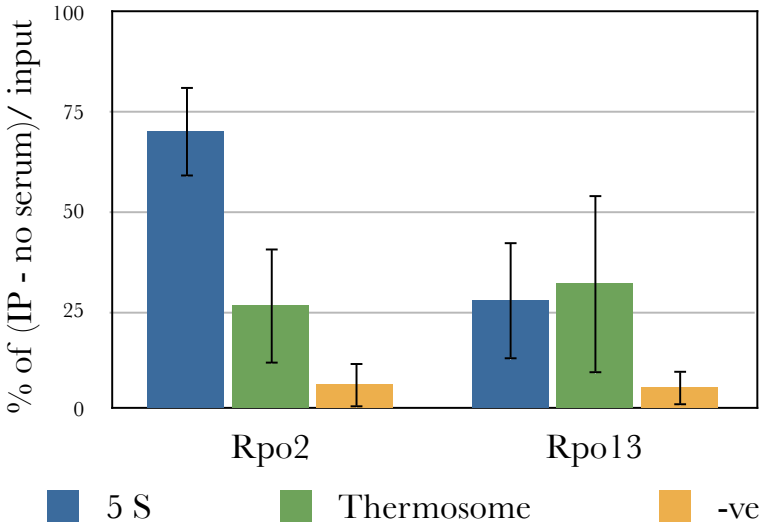


Figure 7.3: *Percentage of immunoprecipitated DNA by the antibodies against Rpo2 and Rpo13. Values were obtained with a Q-PCR experiment with primers designed against the ribosomal 5S rRNA coding region, the thermosome coding region and a region containing no ChIP-seq probe values defined as negative region for both the antibodies against Rpo2 and Rpo13. It is found that the regions corresponding to the ribosomal 5S rRNA coding region and the thermosome coding region are amplified by the designed primers in both the immunoprecipitated protein-DNA complexes by the antibodies against Rpo2 and Rpo13, while little amplification occurs for the negative region.*

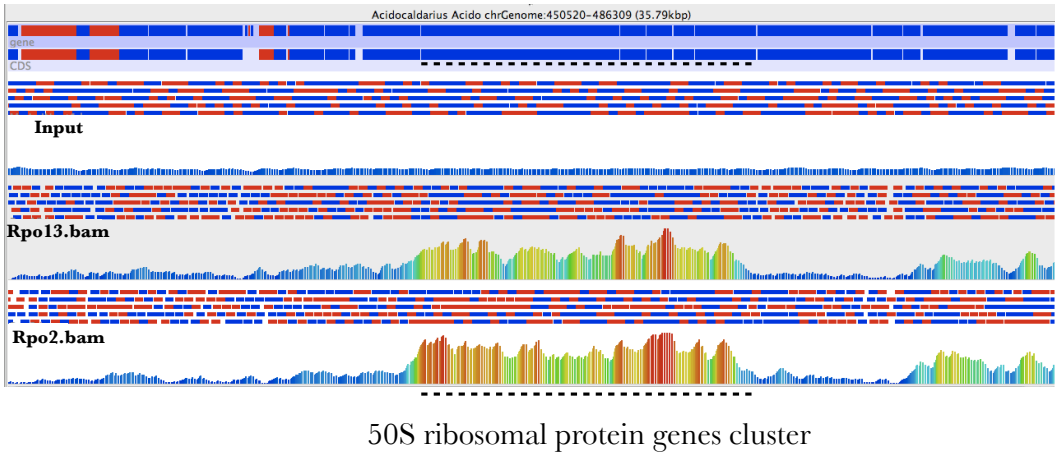


Figure 7.4: *Analysis of the highly transcribed 50S ribosomal protein genes cluster reveals that there is a very good, local correlation between the Rpo2 and Rpo13 subunits. A running window probe generator function with a window size and step size of both 20 bp was applied to the reads.*

but with an additional step to firstly reverse transcribe the RNA into double-stranded cDNA. Subsequent addition of Illumina Indexing adapters on both cDNA

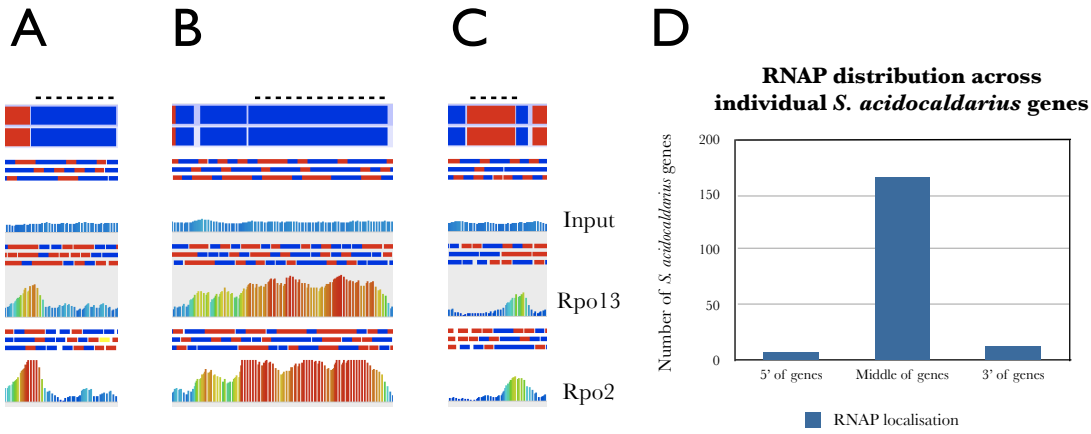


Figure 7.5: *Analysis of the RNAP distribution across a number of individual S. acidocaldarius genes. (**A**) Example of the RNAP defined in terms of Rpo2 and Rpo13 subunits occupying the 5' of a gene. (**B**) Example of the RNAP defined in terms of Rpo2 and Rpo13 subunits occupying the middle of a gene. (**C**) Example of the RNAP defined in terms of Rpo2 and Rpo13 subunits occupying the 3' of a gene. (**D**) In the S. acidocaldarius genome, the RNAP is mainly found in the middle of genes, while occupying either the 5' or 3' end of a subset of genes. A running window probe generator function with a window size and step size of both 20 bp was applied to the Rpo2 and Rpo13 reads in order to analyse the RNAP's distribution.*

ends and 15 cycles of PCR to amplify the amount of prepared material lead to a library with average size of 300 bp. Samples were then sequenced using 55 bp single-end runs with a 7 bp index read. 25,424,885 reads were obtained for the RNA library, with 24,518,376 reads mapped onto the *S. acidocaldarius* genome.

Probe values were assigned to the number of reads mapped to each gene in the *S. acidocaldarius* genome, where contigs closer than 30 bp are merged, the minimum contig size is 250 bp corresponding to the average *S. acidocaldarius* open reading frame, and a value of 0.8 was assigned for a signal to noise coverage defined as depth cutoff. As can be seen in Figure7.7, a total significant number of 1021 probes were obtained and assigned to the number of reads mapped to each gene in the *S. acidocaldarius* genome for both Rpo2 and Rpo13 as well as the number of reads corresponding to levels of abundance of the various *S. acidocaldarius* transcripts. All probe values were standardised to the input sample. In order to correlate RNAP distribution versus transcript abundance, correlation plots were obtained with the RNA-seq and ChIP-seq data sets obtained with the antibodies against Rpo2 and

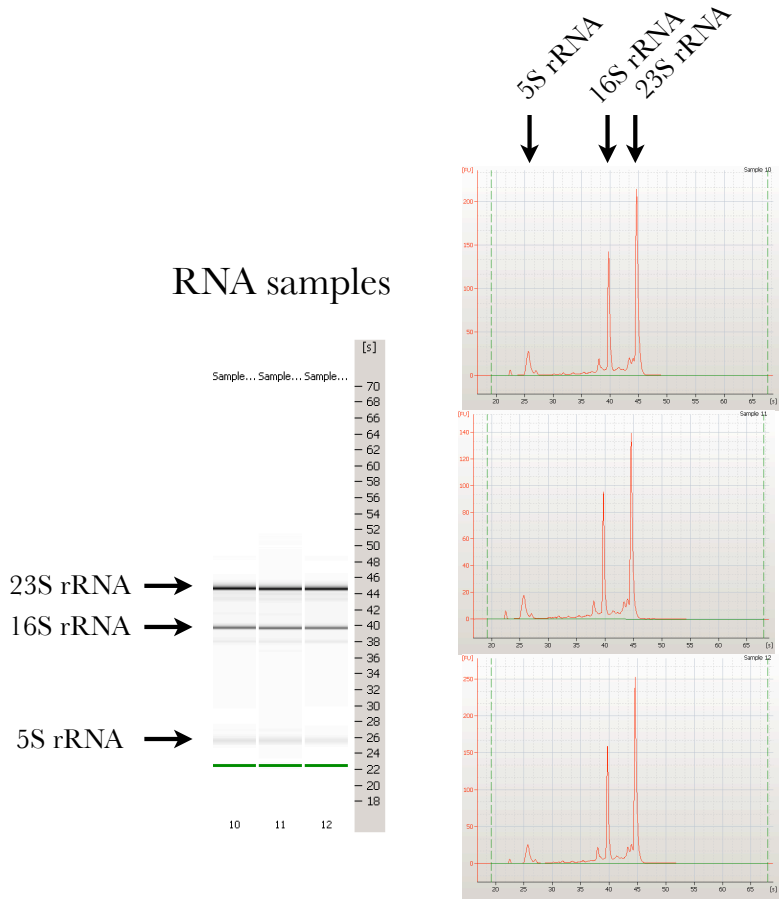


Figure 7.6: *RNA samples prepared from the S. acidocaldarius cell extract, which have been subsequently pooled together for a RNA-seq experiment. The samples are found to contain the predominant 23S, 16S and 5S rRNA species, while tRNAs are under-represented.*

Rpo13 (Figure7.7). As can be seen by the correlation R factor value of around 0.5 for both data sets, there is a global correlation at most genomic regions between transcript abundance and RNAP localisation defined in terms of Rpo2 and Rpo13 subunits (Figure7.7). The 8 genomic regions corresponding to the highest RNAP peaks as well as to highest levels of transcription are listed in Table7.1. However, regions with low RNAP distribution signals but high signals corresponding to high transcript levels are observed as well (Figure7.7). In order to identify transcripts of a higher abundance with poor RNAP localisation, the probes of the RNA-seq were standardised to the probes of the ChIP-seq obtained with the antibody raised against Rpo13 (Figure7.8). The 8 identified coding regions with high levels of transcript

abundance and poor RNAP localisation are listed in Table7.2. Furthermore, it should be noted that paused sites occur, with stalled RNAPs and high levels of transcripts past the RNAP location (Figure7.9). Thus, in summary, while there is an overall, moderately good correlation between RNAP distribution and transcript abundance, high levels of transcripts can occur at coding regions poorly occupied by the RNAP.

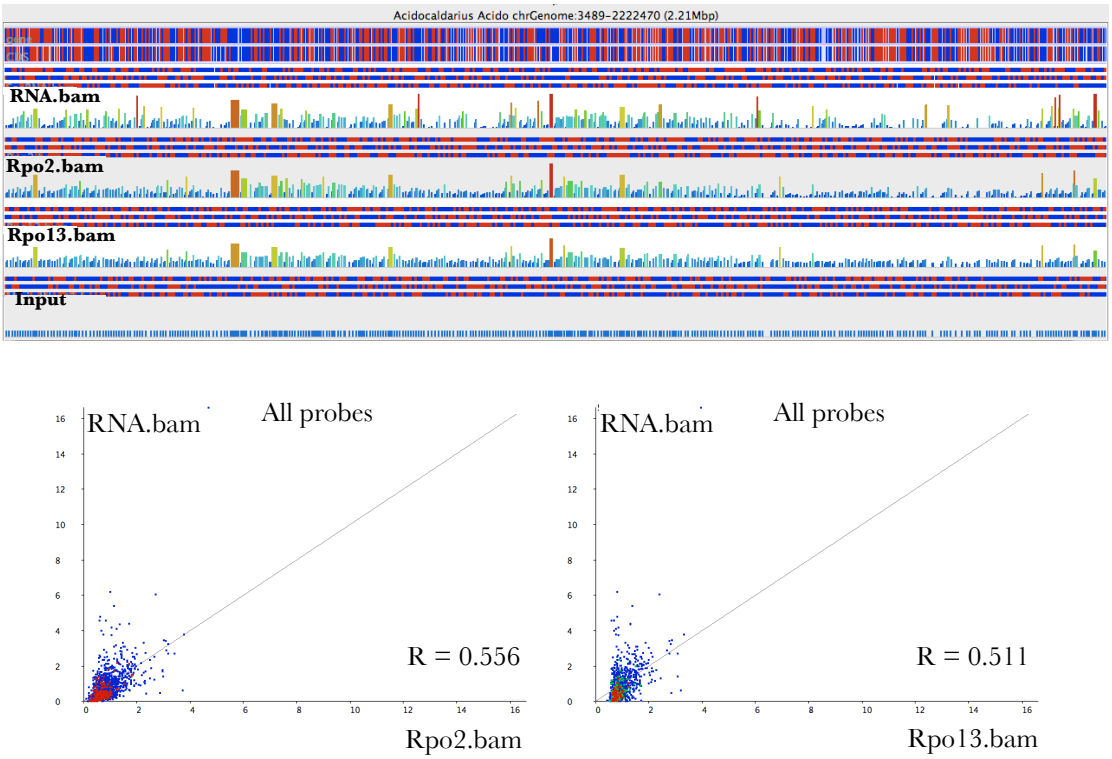


Figure 7.7: *RNA-seq experiment in conjunction with ChIP-seq experiments performed with the antibodies against the RNAP's Rpo2 and Rpo13 subunits. All the values of the obtained 1021 probes for the three data sets are standardised to the input sample. There is a moderately good overall and global correlation at most genomic regions between transcript abundance and RNAP localisation defined in terms of Rpo2 and Rpo13 subunits. A correlation R factor value of 0.5 is obtained upon plotting probes assigned to Rpo2's genomic positions versus transcript abundance, and upon plotting Rpo13's genomic positions versus transcript abundance.*

Peak 1 3.188 2.769 2.694	Peak 2 3.771 3.311 3.799	Peak 3 3.073 2.766 3.412	Peak 4 3.173 2.82 3.256	Peak 5 3.02 2.876 3.477	Peak 6 3.11 2.681 2.894	Peak 7 3.405 3.081 2.716	Peak 8 2.713 2.406 6.064
Saci_0080 (30S ribosomal protein S13P)	Saci_0592 (50S ribosomal protein L22P)	Saci_0694 (DNA-directed RNA polymerase, subunit H)	Saci_0982 (succinate dehydrogenase subunit A)	Saci_1203 (thermosome gamma subunit)	Saci_1455 (alanyl-tRNA synthetase)	Saci_2249 (molybdopterin oxidoreductase)	Saci_2354 (conserved membrane protein)
Saci_0081 (30S ribosomal protein S4P)	Saci_0591 (30S ribosomal protein S3P)	Saci_0693 (DNA directed RNA polymerase subunit B)	Saci_0981 (succinate dehydrogenase subunit B)		Saci_1456 (50S ribosomal protein L12P)	Saci_2250 (pyruvate/ ketoisovalerate oxidoreductase)	
Saci_0082 (30S ribosomal protein S11P)	Saci_0590 (50S ribosomal protein L29P)	Saci_0692 (DNA-directed RNA polymerase subunit A')	Saci_0980 (succinate dehydrogenase subunit C)		Saci_1457 (50S ribosomal protein L10P)	Saci_2251 (pyruvate synthase delta chain)	
Saci_0083 (DNA-directed RNA polymerase subunit D)	Saci_0588 (30S ribosomal protein S17P)	Saci_0691 (DNA-directed RNA polymerase subunit A")	Saci_0979 (succinate dehydrogenase subunit D)		Saci_1458 (50S ribosomal protein L1P)	Saci_2252 (pyruvate synthase alpha chain)	
Saci_0084 (50S ribosomal protein L18E)	Saci_0587 (50S ribosomal protein L14P)	Saci_0690 (50S ribosomal protein L30E)	Saci_0975 (single strand DNA binding protein)		Saci_1459 (50S ribosomal protein L11P)	Saci_2253 (pyruvate synthase beta chain)	
Saci_0085 (50S ribosomal protein L13P)	Saci_0586 (50S ribosomal protein L24AB)	Saci_0688 (30S ribosomal protein S12P)	Saci_0973 (molybdopterin biosynthesis protein)		Saci_1460 (NusG)	Saci_2254 (aldose reductase)	
Saci_0086 (30S ribosomal protein S9P)	Saci_0585 (30S ribosomal protein S4E)	Saci_0686 (30S ribosomal protein S7P)			Saci_1461 (preprotein translocase SecE subunit)		
Saci_0087 (DNA-directed RNA polymerase subunit N)	Saci_0584 (50S ribosomal protein L5P)	Saci_0685 (elongation factor 1-alpha)			Saci_1462 (signal recognition particle)		
Saci_0088 (30S ribosomal protein S2P)	Saci_0583 (30S ribosomal protein S14P)	Saci_0684 (30S ribosomal protein S10P)			Saci_1464 (50S ribosomal protein LX)		
Saci_0091 (isopentenyl-diphosphate delta-isomerase)	Saci_0582 (30S ribosomal protein S8P)				Saci_1465 (translation initiation factor)		

Peak value for Rpo2
Peak value for Rpo13
Peak value for RNA

Table 7.1: *Coding regions corresponding to the highest RNAP peaks and highest levels of transcription. These coding regions code for the 30S ribosomal proteins, 50S ribosomal proteins, RNAP subunits, metabolic enzymes, translation factors, the ssDNA binding protein, the thermosome γ subunit, tRNA synthetases, the elongation factor NusG, translocase subunits, the signal recognition particle protein and membrane proteins. Peak values for RNAP localisation and RNA expression levels are detailed for each peak (red for Rpo2 peaks, blue for Rpo13 peaks and green for RNA peaks).*

7.2.3 Origin Firing versus Transcript Abundance and RNAP Localisation

The input sample represents the DNA content of exponentially growing *S. acidocaldarius* cells. As cells grow and divide, DNA replication of the chromosome also occurs, where a short G1 period between cell division and DNA replication is thought to occur [149]. The input sample contains DNA recovered from

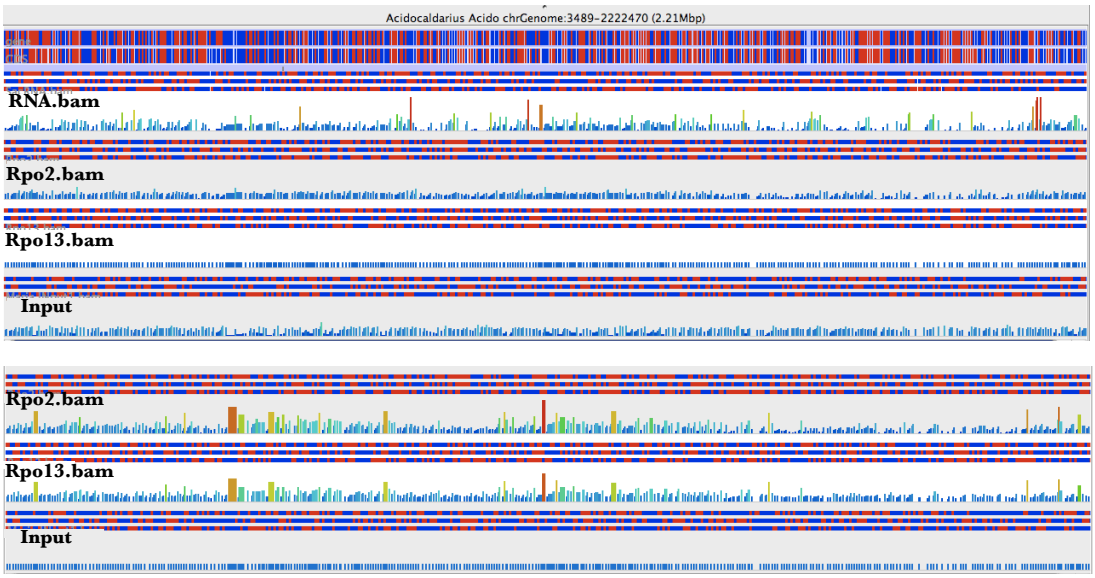


Figure 7.8: *RNA-seq experiment in conjunction with ChIP-seq experiments performed with the antibodies against the RNAP's Rpo2 and Rpo13 subunits. On the top pannel, all the values of the obtained 1021 probes for the three data sets are standardised to the Rpo13 sample. On the lower pannel, all the values of the obtained 1021 probes for the two data sets are standardised to the input sample. Some coding regions have high expression levels with high transcript probe values and poor RNAP occupancy, where they can be individuated by their corresponding emerging RNA peaks (top pannel).*

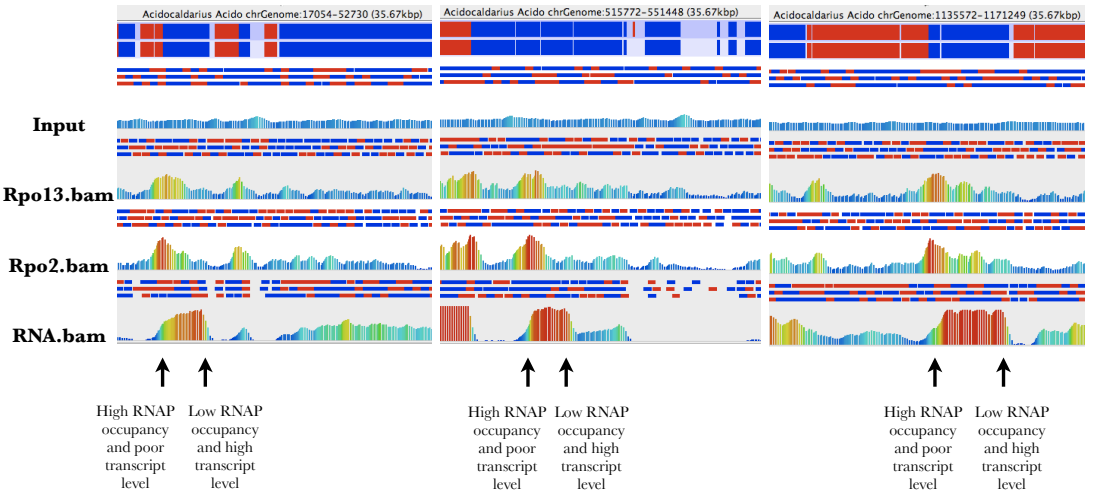


Figure 7.9: *Genomic loci where paused sites occur, where high RNAP occupancy and poor transcript levels is followed by low RNAP occupancy and high transcript levels. This indicates stalled RNAPs. A running window probe generator function with a window size and step size of both 20 bp was applied to the reads.*

exponentially growing *S. acidocaldarius* cells, which is asynchronous and therefore contains cells at different stages of DNA replication and cell division. Cells have DNA

Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8
3.354 1.707 1.307 4.382	3.85 0.796 0.65 2.502	5.887 0.63 0.814 4.792	4.972 0.754 0.758 3.771	3.903 1.149 1.386 5.411	3.91 1.154 0.79 3.088	5.67 1.25 0.758 4.11	6.527 0.593 0.704 4.593
Saci_0315 (electron transfer flavoprotein)	Saci_0763 (thiamine pyrophosphate-requiring enzyme)	Saci_1037 (peptide transport system permease protein)	Saci_1265 (succinyl-CoA synthetase beta chain)	Saci_1301 (ribose 5-phosphate isomerase)	Saci_2269 (carbon monoxide dehydrogenase medium chain)	Saci_2277 (conserved Archaeal membrane protein)	Saci_2282 (transporter protein)
Saci_0316 (FixC)		Saci_1038 (conserved protein)			Saci_2270 (carbon monoxide dehydrogenase small chain)		
					Saci_2271 (carbon monoxide dehydrogenase large chain)		

- **Peak value for RNA upon standardisation to Rpo13**
- **Peak value for Rpo2 upon standardisation to input**
- **Peak value for Rpo13 upon standardisation to input**
- **Peak value for RNA upon standardisation to input**

Table 7.2: *Coding regions corresponding to the highest levels of transcription with poor RNAP occupancy. These coding regions code for electron transfer proteins, metabolic enzymes, peptide transport proteins and membrane proteins. Peak values for RNAP localisation and RNA expression levels are detailed for each peak (red for RNA peaks standardised to Rpo13 signals, blue for Rpo2 peaks standardised to input sample, green for Rpo13 peaks standardised to input sample and violet for RNA peaks standardised to input sample).*

replication origins moving across the entire genome at different DNA replication stages. The input DNA sample from these cells also contains an average of replicated and non-replicated DNA regions across the entire chromosome, and thus the DNA content in terms of n or 2n copies. Therefore, the analysis of DNA replication in the input sample will reveal an average of the position of DNA replication across the genome. This concept is defined as marker frequency analysis, with the frequency of a marker as a function of genomic position. Since the positions of *S. acidocaldarius* origin firing are already known [51], no standardisation with respect to that of cells in stationary phase is required, and a moving window average function can be directly applied to the input sample to smoothen the signal to noise ratio and produce a DNA profile in terms of replicated regions. As can be observed in (Figure7.10), three peaks corresponding to the positions of origin firing were obtained, with *oriC2* at position 0 bp, *oriC1* at position 579,000 bp, and *oriC3* at position 1197,000

bp. When the peaks corresponding to origin firing were correlated with the RNAP distribution and transcript abundance, it is found that the positions of origin firing are in the vicinity of RNAP peaks as well as transcript level peaks, some of which correspond to genomic regions that code for proteins involved in DNA replication coupled with recombination, such as the DNA recombination protein RadA. Thus, these genomic regions may code for proteins required for the overall process of origin firing and DNA replication. In conclusion, the two processes of transcription and DNA replication may be concerted and form one major complex process.

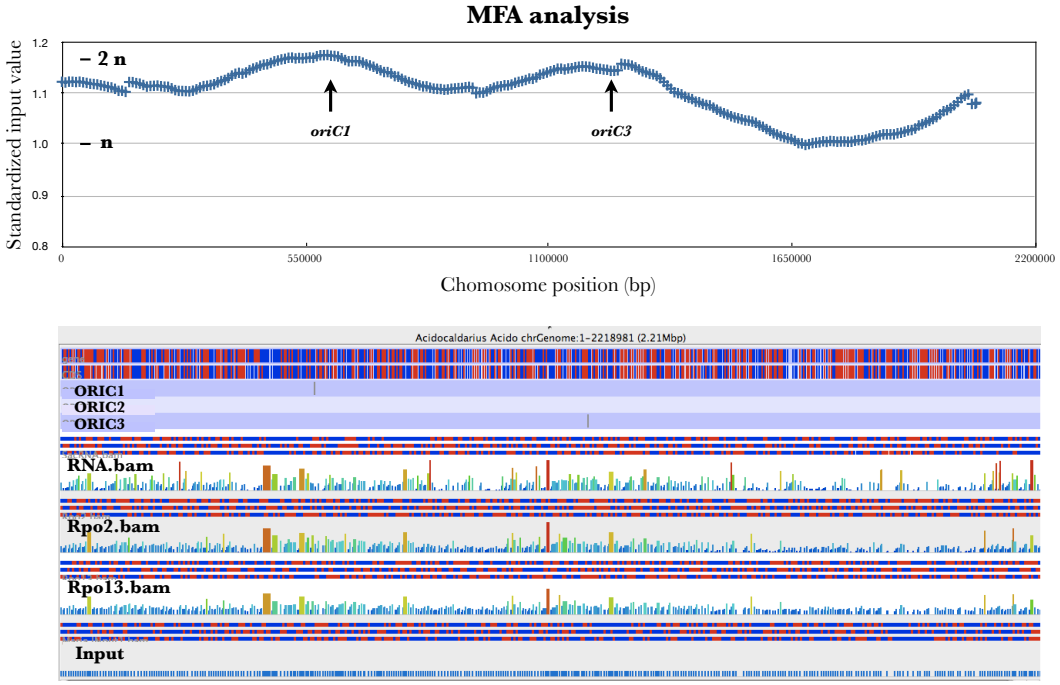


Figure 7.10: *MFA analysis of the input sample obtained from exponentially growing S. acidocaldarius cells. Since the positions of origin firing are known (oriC2 at position 0 bp, oriC1 at position 579,000 bp, and oriC3 at position 1197,000 bp), no standardisation with respect to that of cells in stationary phase is required. Upon the generation of a moving window average function applied to input sample values, the DNA profile defined in terms of replicated regions is obtained. When overlapped with the profile of RNAP occupancy versus the one corresponding to transcript level abundance, some RNAP and transcript peaks are in the vicinity of origin firing. Their corresponding genomic regions code for proteins that may be potentially involved in the process of DNA replication, leading to the coupling of transcription with DNA replication.*

7.3 Discussion

ChIP-seq with antibodies against Rpo2 and Rpo13 as well as RNA-seq experiments have been performed in order to analyse the distribution of the RNAP in terms of Rpo2 and Rpo13 subunits in conjunction with the analysis of transcript expression levels. Indeed, RNAP occupancy within the gene is defined by RNAP recruitment to the promoter, open complex formation and promoter escape, the rate of each step being dependent on promoter sequence and subject to regulation by transcriptional regulators, which resemble to bacterial regulators in terms of their *modus operandi* [13], [16]. The ChIP-seq procedure with the antibodies against Rpo2 and Rpo13 enabled the analysis of the archaeal RNAP transition from transcriptional initiation to elongation *in vivo*, since it determines the RNAP density at promoters and their corresponding transcribed regions. If the transition is fast, RNAP levels should be constant across the transcribed regions, while if the transition is slow, higher levels of RNAP should be present at promoter regions rather than in the coding regions. Studies of the RNAP genomic occupancy in bacteria and eukaryotes have revealed a multitude of RNAP binding patterns in the genome [141], [124], [76], [49], [166], [84], [143], [34], [40], [99], [111], [145], [153], [167]. Specifically, ChIP-chip analyses of the *E. coli* RNAP distribution in the genome showed a higher RNAP association at promoter regions than within coding sequences for most, but not all transcription units ([141], [124], [76], [49], [166]), while with other bacterial RNAPs, such as the *Bacillus subtilis* RNAP, the RNAP is evenly distributed from the promoter to the coding sequences [84]. The *E. coli* RNAP distribution across transcribed regions was found to be gene dependent, where the RNAP was also found to stall at multiple promoters with no detectable transcriptional activity, with post-recruitment steps of transcription initiation being often rate limiting and the extent to which they are rate limiting varying dramatically upon the promoter [141], [76]. Furthermore, a more in depth investigation revealed that promoter-proximal RNAP peaks are offset from the σ factor, but rather co-occur with peaks corresponding to the elongation

factor NusA and the termination factor ρ, suggesting that the RNAP peaks reflect elongating complexes, rather than initiating complexes [124]. However, inhibition of ρ did not increase RNAP levels within genes downstream from the RNAP peaks, suggesting that the peaks are caused by transcriptional attenuation caused by a mechanism other than ρ-dependent attenuation. This may differ from the other types of mechanisms such as promoter trapping and promoter-proximal pausing [124]. Moreover, bacterial RNAP occupancy was also found to be dependent on different growth conditions, with the RNAP occupying distinct binding sites that are within transcribed coding regions upon rifampicin treatment. These are not normally occupied by the RNAP, indicating promoter-competent regions [49]. A similar situation occurs in *Drosophila* and mammalian cells ([145], [99], [40], [34]), where paused RNAP complexes in early elongation are dependent on the TATA box and initiator sequences. These have been characterised together with promoter-localised RNAP-II at numerous locations such as developmentally regulated genes. A different situation occurs in *S. cerevisiae* [111], [153], where RNAP levels are constant across the coding regions, indicating that the transition between transcriptional initiation and elongation is fast and differently controlled. Apparently, the pre-initiation complex formation is the major rate-limiting step for transcription [111], [153]. Promoter-localised RNAP-II is only observed in stationary phase *S. cerevisiae* [140]. The analysis of the archaeal RNAP distribution defined in terms of Rpo2 and Rpo13 subunits complements all the previous RNAP analyses. It is found that some genomic regions are highly occupied by the archaeal RNAP, while other regions are only poorly occupied. Here the RNAP appears to display an oscillatory pattern of gene occupation for both Rpo2 and Rpo13. Furthermore, the RNAP does not occupy the coding regions with any preference for either the 5' end or the 3' end of genes, but rather the middle of the gene. Therefore, the RNAP distribution across the coding regions suggests that the transition from initiation to elongation is fast as with the *S. cerevisiae* case, where either RNAP recruitment or pre-initiation complex formation are rate-limiting. This conclusion is based on the assumption that RNAP

crosslinking efficiency is similar at all stages of the transcription cycle which is highly likely since RNAP is in close proximity to large regions of DNA throughout the transcription cycle. In summary, there are significant species-specific differences in rate-limiting steps during the transition from initiation to elongation. This can lead to either a promoter-proximal distribution of the RNAP or a constant one across the coding regions such as in Archaea.

My analysis of gel filtration elution volumes of Rpo13 versus the RNAP in the *S. acidocaldarius* cell extract, suggested that Rpo13 might behave more as a dissociable transcription factor as opposed to a permanent and *bona fide* RNAP subunit. This in turn leads to the existence of a free Rpo13 pool within the cell. Analysis of the ChIP-seq data obtained with the Rpo2 and Rpo13 antibodies has however revealed a close correlation between Rpo2 and Rpo13 subunits within the *S. acidocaldarius* genome, where a correlation R factor value of 0.9 is obtained. This suggests that Rpo13 is associated with the bulk RNAP at all stages of transcription, and does not dissociate from the RNAP after transcription initiation. Therefore, Rpo13 appears to be a true RNAP subunit, which contributes to the formation of a functional RNAP throughout the transcription cycle. The free Rpo13 pool in the cell extract might reflect non-functional and unincorporated Rpo13 following RNAP assembly. At initiation, Rpo13 may help to unwind DNA through DNA contacts with its N- and C-termini either while the open-complex is being formed, or soon after. Indeed, in the RNAP-DNA crystal structure, contacts between the Rpo13 N- and C-termini and the bubble oligo are observed straight after the bubble region ([180]). Upon promoter escape, Rpo13 could contribute to the overall RNAP structural plasticity required for RNAP translocation in synergy with Rpo5 and Rpo1 subunits. Specifically, the α1 and α2 helices might perform some structural roles, acting as constitutive anchors onto the RNAP with contacts between the α1 helix and Rpo1 and Rpo5 occur ([180]), and connecting the C-terminal DNA-binding region of Rpo13 to the rest of the RNAP. Rpo13 would therefore have an equivalent role to the Rpb5 jaw domain, absent in Archaea, by gripping the DNA throughout the transcription cycle

with a weak DNA binding affinity of K_d of around 2 μM. This would allow for transcriptional processivity whilst modulating the balance between non-specific and specific DNA interactions. In summary, Rpo13 is a *bona fide* RNAP subunit, where it may contribute to the formation of the open complex at initiation through DNA unwinding. During elongation, it may help the RNAP to grip onto the DNA and so contribute to its overall structural plasticity through contacts with Rpo1 and Rpo5 needed for the translocation process.

The ChIP-seq procedure performed with the antibodies against Rpo2 and Rpo13 was performed in conjunction with an RNA-seq of the *S. acidocaldarius* transcripts in order to correlate RNAP occupancy within the genome with the various levels of transcript expression. It is found that some genomic regions have a high level of RNAP abundance in conjunction with a high expression level of transcripts, while other regions have a low level of RNAP abundance in conjunction with a high expression level of transcripts and finally some regions display high RNAP levels with low transcript abundance. A total number of 8 high RNAP peaks corresponding to high levels of transcription are observed. The corresponding coding regions include regions that code for the 30S ribosomal proteins, 50S ribosomal proteins, RNAP subunits, metabolic enzymes, translation factors, the ssDNA binding protein, the thermosome γ subunit, tRNA synthetases, the elongation factor NusG, translocase subunits, the signal recognition particle protein and membrane proteins. Indeed, these proteins and enzymes have a pivotal role in the functioning for the cell, and a high abundance due to high expression levels is to be expected. Moreover, high protein levels due to low rates of mRNA decay also occur. The 8 identified highly stable mRNA transcripts include transcripts that code for electron transfer proteins, metabolic enzymes, peptide transport proteins and membrane proteins. All of these proteins play a fundamental role in the cell. The quantification of the rate of mRNA decay by an independent method consisting in the measuring of transcript abundance in samples collected at different time points after transcriptional arrest by actinomycin D has been performed [4].

This analysis revealed that transcripts encoding proteins involved in growth-related processes such as transcription, translation and central metabolic pathways generally decayed rapidly, while those encoding products necessary for maintaining cellular homeostasis such as ion transporters are more stable [4]. More specifically, short half-lives were defined by a $t_{1/2}$ value between 0 and 4, while long half-lives were defined by a $t_{1/2}$ value of more than 9 [4]. The $t_{1/2}$ values for the transcripts corresponding to the identified coding regions with high expression levels and poor RNAP occupancy are listed in (Table7.3), and are in good correlation with the high observed RNA peaks. Furthermore, sites corresponding to paused RNAPs with higher transcript levels further down the stalled RNAPs are also observed upon the analysis of the *S. acidocaldarius* RNAP distribution versus transcript level expression. Therefore, the RNAP generally occupies coding regions with preference for the middle of genes, where RNAP pausing can occur. Regulation at the level of transition from initiation to elongation or at the level of elongation is most likely due to the action of archaeal regulators [16], [132]. Finally, coding regions located near the origins of replication *oriC1*, *oriC2* and *oriC3* have high RNAP occupancy coupled with high levels of transcription. Some of these regions code for proteins involved in DNA replication coupled with recombination such as the DNA recombination protein RadA. Thus, the two processes of DNA replication and transcription may be coordinated in a spatial and temporal manner. Upon transcription of these proteins, the vicinity of the origins facilitates their *modus operandi* at the origins, resulting in an overall concerted process. In summary, some transcripts associated to coding regions important for the cell functioning and homeostasis occur at high levels upon RNA-seq analysis, either in a dependent or independent fashion from the levels of RNAP occupancy.

In conclusion, ChIP-seq analyses performed with the antibodies against Rpo2 and Rpo13 in conjunction with a RNA-seq to analyse the levels of transcript expression have revealed novel insights in the pattern of RNAP's distribution in the genome, which appears to be constant throughout the coding genes with no

Peak 1 3.354 1.707 1.307 4.382	**Peak 2** 3.85 0.796 0.65 2.502	**Peak 3** 5.887 0.63 0.814 4.792	**Peak 4** 4.972 0.754 0.758 3.771	**Peak 5** 3.903 1.149 1.386 5.411	**Peak 6** 3.91 1.154 0.79 3.088	**Peak 7** 5.67 1.25 0.758 4.11	**Peak 8** 6.527 0.593 0.704 4.593
Saci_0315 (electron transfer flavoprotein) $t_{1/2}$ = 4.8 $t_{1/2}$ = 4.48	Saci_0763 (thiamine pyrophosphate-requiring enzyme) Not avaliable	Saci_1037 (peptide transport system permease protein) $t_{1/2}$ = 3.56 $t_{1/2}$ = 3.84	Saci_1265 (succinyl-CoA synthetase beta chain) $t_{1/2}$ = 4.34 $t_{1/2}$ = 4.09	Saci_1301 (ribose 5-phosphate isomerase) $t_{1/2}$ = 3.93 $t_{1/2}$ = 3.67	Saci_2269 (carbon monoxide dehydrogenase medium chain) $t_{1/2}$ = 3.62	Saci_2277 (membrane protein) $t_{1/2}$ = 5.25 $t_{1/2}$ = 5.50	Saci_2282 (transporter protein) $t_{1/2}$ = 4.88 $t_{1/2}$ = 4.01
Saci_0316 (FixC) $t_{1/2}$ = 4.61 $t_{1/2}$ = 4.10					Saci_2270 (carbon monoxide dehydrogenase small chain) Not avaliable		
					Saci_2271 (carbon monoxide dehydrogenase large chain) $t_{1/2}$ = 4.28		

- **Peak value for RNA upon standardisation to Rpo13**
- **Peak value for Rpo2 upon standardisation to input**
- **Peak value for Rpo13 upon standardisation to input**
- **Peak value for RNA upon standardisation to input**

Table 7.3: *Coding regions corresponding to the highest levels of transcription with poor RNAP occupancy. These coding regions code for electron transfer proteins, metabolic enzymes, peptide transport proteins and membrane proteins. Their corresponding $t_{1/2}$ values are indicated in red ([4]), where short half-lives are defined by a $t_{1/2}$ value between 0 and 4, while long half-lives are defined by a $t_{1/2}$ value of more than 9.*

preference for promoter regions, with high peak levels at genes involved in the growth of the cell such as transcription, translation and metabolic pathways. Furthermore, Rpo13 is found to be a *bona fide* RNAP subunit. A possible role at transcription initiation could be DNA unwinding, while at elongation, Rpo13 may help in DNA gripping and overall RNAP structural plasticity. Moreover, some mRNA transcripts coding for proteins involved in general cellular homeostasis were found to be highly stable, where their coding regions are poorly occupied by the RNAP. In addition, a concerted process involving origin firing and transcription may occur. Finally, the single archaeal RNAP might not be defined by a regulatory role of RPC34 neither at specific genes, such as non-coding genes or functionally related genes, nor at a general level, where RPC34 could act as a non-specific, general transcriptional regulator (data not shown).

Chapter 8

Conclusions

The single archaeal RNAP is similar to the eukaryotic RNAP-II in terms of subunit composition and overall protein structure. Specifically, it contains 13 subunits, as opposed to 12 subunits for the RNAP-II. Despite its similarity, the archaeal RNAP has some unique differences to RNAP-II. First, it does not contain a homologue of the eukaryotic-specific Rpb9 subunit. Secondly, it lacks the Rpb1 C-terminal domain, which in eukaryotes, is composed of up to 52 heptapeptide repeats (YSPTSPS). These repeats are sites for phosphorylation critical for the transition from the initiation phase to processive elongation of the transcript [81], [75]. Thirdly, and most importantly, the archaeal Rpo5 subunit differs from the eukaryotic Rbp5 equivalent in that it does not have the Rpb5 N-terminal module but only the Rpb5 C-terminal globe. The space occupied by the Rpb5 N-terminal module in RNAP-II, which is part of the RNAP-II lower jaw module and is involved in downstream DNA gripping, is replaced by a newly identified, archaeal-specific subunit denoted as Rpo13 [96]. Because Rpo13 is highly basic, where its C-terminal tail contains conserved lysine residues, and because its position is analogous to the jaw domain of Rbp5, it is proposed that Rpo13 binds to DNA in a similar manner to the jaw domain of Rpb5. Furthermore, an archaeal homologue of the eukaryotic, RNAP-III-specific RPC34 subunit has been identified [22]. Upon analysis of the genomic context of the gene coding for the archaeal RPC34, it was found that the *rpc34* gene co-

localises with genes coding for factors involved in the transcription, modification and processing of structural rRNAs and tRNAs, such as a tRNA pyrophosphorylase, and a paralog of TFB referred to as TFB-2 [22]. Therefore, it is proposed that RPC34 interacts with the single archaeal RNAP, where RPC34's temporary and transient association with the archaeal RNAP defines an evolutionary precursor of RNAP-III, the latter enzyme being permanently associated with RPC34. More importantly, RPC34 could contribute to the regulation of transcription of the archaeal RNAP, re-directing it to a subset of genes in complex with TFB-2, such as non-coding rRNA and tRNA genes. RPC34 may also act at the DNA level, promoting open complex formation [29].

The Rpo13 and RPC34 proteins were successfully purified by recombinant DNA techniques. Rpo13 is a novel archaeal RNAP subunit and RPC34 is a potential RNAP interactant partner. Polyclonal antibodies were successfully raised against recombinant, native Rpo13 and denatured RPC34. The latter antibody also recognises endogenous RPC34 in *S. acidocaldarius* cell extracts. Biochemical analysis of Rpo13 by means of EMSA assays revealed that Rpo13 associates with the DNA with a binding affinity of K_d of around 2 μM. The identity of Rpo13 in the protein-DNA complexes was verified with a supershift EMSA assay with the antibody raised against Rpo13, as well as EMSA assays coupled to an independent transfer of DNA to a nylon membrane and protein to a nitrocellulose membrane. This procedure revealed co-localisation of the two signals. Purification of three Rpo13 mutants (ΔN Rpo13, ΔC Rpo13 and core Rpo13) showed that, upon EMSA analysis, the C-terminal tail of Rpo13 principally mediates DNA binding, since DNA binding comparable to wild-type levels is observed for the ΔN Rpo13 mutant, but not for the ΔC and core Rpo13 mutants. Furthermore, by means of major groove or minor groove competition assays, it is found that Rpo13 associates not with the DNA grooves, but most likely with the negatively charged phosphate backbone via its positively charged C-terminal tail. Moreover, Rpo13 associates with the DNA without introducing any torsional stress that induces DNA supercoiling. In

addition, a gel filtration analysis that compares the elution volumes of endogenous, *S. acidocaldarius* cell extract Rpo13 versus recombinant Rpo13 reveals that Rpo13 in *S. acidocaldarius* cell extract is mainly associated with the bulk RNAP, while a free Rpo13 pool elutes at an elution volume corresponding to recombinant Rpo13. This suggests the existence of RNAP-free Rpo13 protein in *S. acidocaldarius*.

My suggested finding that RNAP-free Rpo13 protein exists in the cell extract that associates with the RNAP only transiently, where it could exert a functional role analogous to the eukaryotic initiation factors absent in Archaea. Thus, Rpo13 could be associated with the RNAP only at transcription initiation, leading to the free Rpo13 pool upon promoter escape of the RNAP. This is further supported by the fact that Rpo13 forms dimers in solution like TFIIF [179], and shows some similarity to the charged region of Tfg1-like factors of TFIIF [180]. Indeed, only TBP and TFB are required for efficient, *in vitro* archaeal transcription [77], [138]. Crystallisation of the archaeal RNAP in conjunction with a TBP-TFB fusion construct, shown in this study to bind to a T6 promoter sequence containing the TATA-like A box as well as to be able to direct *in vitro* transcription, will yield novel insights into the mode of recruitment of the archaeal RNAP by TBP and TFB onto the DNA. Furthermore, I propose that, in conjunction with other RNAP subunits, Rpo13 may contribute to open-complex formation through Rpo13's contacts with the DNA, which causes DNA unwinding by pulling the DNA strands. Rpo13 could therefore replace some of the functions exerted by transcription initiation factors in transcription initiation. However, no *in vitro* transcription product arising from the internal TSS of a bubble DNA oligo substrate, shown to make contacts with the E32 residue of the Rpo13's N-terminus and proposed to interact with the C-terminal tail of Rpo13 ([180]), is observed. The removal of Rpo13 from RNAP by alternative methods may allow a more direct role of Rpo13 in transcription to be assessed.

RPC34's interaction with the RNAP and TFB-2 was tested in this study with either *in vitro* or *in vivo* means. It is found that while a GST-based pull down assay suggests the existence of an interaction between the single, archaeal RNAP and

RPC34, such an interaction cannot be detected *in vivo* with co-immunoprecipitation assays. It is possible that the interaction is weak, or only a small fraction of RNAP in the cell is associated with RPC34, or that the interaction is cell cycle regulated. However, upon 'baby machine' *S. acidocaldarius* cell cycle synchronisation and analysis of RPC34 protein levels throughout the cell cycle, no fluctuations in the RPC34 protein expression levels were observed when compared to the TBP protein used as a loading control. Furthermore, no interaction between RPC34 and TFB-2 was observed when tested either with a yeast two-hybrid screen or by co-expression means. It is possible that the interaction is weak, transient or additional components are required to bridge such interaction. More importantly, since RPC34 has been observed to cross-link to DNA ([29], [11]), RPC34 could act at the DNA level rather than at the protein level, with a DNA regulatory role as a transcription factor.

I finally undertook a ChIP-seq approach with the antibodies against Rpo2 and Rpo13 subunits since my biochemical studies suggest that Rpo13 has a transcription factor-like role, with only transient association with the archaeal RNAP leading to the existence of a free Rpo13 pool in the *S. acidocaldarius* cell extract. This ChIP-seq analysis of the genomic distribution of the Rpo2 and Rpo13 subunits in the *S. acidocaldarius* genome allowed the correlation of the distribution of Rpo2 versus the one of Rpo13, testing whether Rpo13 is a *bona fide* RNAP subunit, or associates with the RNAP only at transcription initiation in a transcription factor-like mode. The ChIP-seq analyses were performed in conjunction with a RNA-seq analysis to compare global transcription.

Analysis of the genomic distribution of Rpo13 versus the one of Rpo2 revealed that there is a very good correlation between Rpo13 and Rpo2 subunits at both a global and local level. Therefore, Rpo13 behaves as a *bona fide* RNAP subunit, since it associates with the RNAP not only at transcription initiation, but also during transcription elongation. Furthermore, the genomic distributions of Rpo13 and Rpo2 across the *S. acidocaldarius* genome are relatively even, suggesting that the transition from transcription initiation to elongation is fast in Archaea, similarly

to *S. cerevisiae* [111], [153]. Moreover, upon RNA-seq analysis, I found that the levels of transcription at some genomic loci is higher than at the rest of the loci, independent of RNAP occupancy levels. Thus, Rpo13 is a *bona fide* RNAP subunit, with a functional role at both transcription initiation and elongation. At transcription initiation, Rpo13 could contribute to open-complex formation through DNA unwinding by pulling the DNA strands, while during transcription elongation, it could contribute to the overall RNAP plasticity required for RNAP translocation while gripping the DNA in similarity to the Rpb5 jaw domain missing in Archaea. The weak Rpo13's binding affinity to the DNA of K_d of around 2 μM allows for transcriptional processivity.

In conclusion, this study has contributed to the advancement of knowledge surrounding RPC34 and Rpo13. It argues against a role for RPC34 in the RNAP recruitment for the transcription and regulation of a subset of genes such as non-coding rRNA and tRNA genes similarly to the eukaryotic RNAP-III-RPC34. However, it supports a role for the archaeal-specific Rpo13 subunit in the transcription of all the genes throughout the entire RNAP transcription cycle. At initiation, Rpo13 could promote open-complex formation through DNA unwinding, while at elongation, it could grip DNA and contribute to the overall RNAP plasticity. My biochemical *in vitro* analyses in conjunction with *in vivo* approaches such as co-immunoprecipitation and ChIP-sequencing assays can in the future be complemented with gene knock-outs of the *rpo13* and *rpc34* genes to assess any direct roles of Rpo13 and RPC34 in transcription. The manipulation of archaeal genetics is however hampered by the scarcity of archaeal antibiotics. Alternative selection methods based on auxotrophic markers are being currently developed and may open up Archaea to molecular genetic analysis [3].

Bibliography

[1] Aagaard, C., Dalgaard, J. Z., and Garrett, R. A. Inter-cellular mobility and homing of an archaeal rDNA intron confers selective advantage over intron-cells of *Sulfolobus acidocaldarius*. *Proc. Natl. Acad. Sci. USA 92* (1995), 12285–12289.

[2] Achenbach-Richter, L., Gupta, R., Zillig, W., and Woese, C. Rooting the archaebacterial tree: the pivotal role of *Thermococcus celer* in archaebacterial evolution. *Syst. Appl. Microbiol. 10* (1988), 231–240.

[3] Allers, T., and Mevarech, M. Archaeal genetics - the third way. *Nat. Rev. Genet. 6*, 1 (2005), 58–73.

[4] Andersson, A. F., Lundgren, M., Eriksson, S., Rosenlund, M., Bernander, R., and Nilsson, P. Global analysis of mRNA stability in the archaeon *Sulfolobus*. *Genome Biol. 7*, 10 (2006), R99.

[5] Andrau, J. C., Sentenac, A., and Werner, M. Mutagenesis of yeast TFIIIB70 reveals C-terminal residues critical for interaction with TBP and C34. *J. Mol. Biol. 288*, 4 (1999), 511–520.

[6] Andrecka, J., Lewis, R., Bruckner, F., Lehmann, E., Cramer, P., and Michaelis, J. Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc. Natl. Acad. Sci. 105*, 1 (2008), 135–140.

[7] Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M., and Iyer, L. M. The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev. 29*, 2 (2005), 231–262.

[8] Armache, K.-J., Kettenberger, H., and Cramer, P. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl. Acad. Sci. 100*, 12 (2003), 6964–6968.

[9] Asturias, F. J., Meredith, G. D., Poglitsch, C. L., and Kornberg, R. D. Two conformations of RNA polymerase II revealed by electron crystallography. *J. Mol. Biol. 272*, 4 (1997), 536–540.

[10] Baliga, N. S., Goo, Y. A., Ng, W. V., Hood, L., Daniels, C. J., and DasSarma, S. Is gene expression in halobacterium NRC-1 regulated by multiple TBP and TFB transcription factors? *Mol Microbiol 36*, 5 (2000), 1184–1185.

[11] Bartholomew, B., Durkovich, D., Kassavetis, G. A., and Geiduschek, E. P. Orientation and topography of RNA polymerase III in transcription complexes. *Mol. Cell. Biol. 13*, 2 (1993), 942–952.

[12] Bell, S. D., Brinkman, A. B., van der Oost, J., and Jackson, S. P. The archaeal TFIIEalpha homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep. 2*, 2 (2001), 133–138.

[13] Bell, S. D., and Jackson, S. P. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends. Microbiol. 6*, 6 (1998), 222–228.

[14] Bell, S. D., and Jackson, S. P. Charting a course through RNA polymerase. *Nat. Struct. Biol. 7*, 9 (2000), 703–705.

[15] Bell, S. D., and Jackson, S. P. The role of transcription factor B in transcription initiation and promoter clearance in the archaeon *Sulfolobus acidocaldarius*. *J. Biol. Chem. 275*, 17 (2000), 12934–12940.

[16] Bell, S. D., and Jackson, S. P. Mechanism and regulation of transcription in Archaea. *Curr. Opin. Microbiol. 4*, 2 (2001), 208–213.

[17] Bell, S. D., Jaxel, C., Nadal, M., Kosa, P. F., and Jackson, S. P. Temperature, template topology, and factor requirements of archaeal transcription. *Proc. Natl. Acad. Sci. 95*, 26 (1998), 15218–15222.

[18] Bell, S. D., Kosa, P. L., Sigler, P. B., and Jackson, S. P. Orientation of the transcription preinitiation complex in Archaea. *Proc. Natl. Acad. Sci. 96*, 24 (1999), 13662–13667.

[19] Bernander, R. The archaeal cell cycle: current issues. *Mol. Microbiol. 48*, 599-604 (2003).

[20] Bernander, R., Poplawski, A., and Grogan, D. W. Altered patterns of cellular growth, morphology, replication and division in conditional-lethal mutants of the thermophilic archaeon *Sulfolobus acidocaldarius*. *Microbiology 146*, 749-757 (2000).

[21] Bischler, N., Brino, L., Carles, C., Riva, M., Tschochner, H., Mallouh, V., and Schultz, P. Localization of the yeast RNA polymerase I-specific subunits. *EMBO J. 21*, 15 (2002), 4136–4144.

[22] Blombach, F., Makarova, K. S., Marrero, J., Siebers, B., Koonin, E. V., and van der Oost, J. Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in archaea. *Biol. Direct. 4* (2009), 39.

[23] Botting, C. H., Talbot, P., Paytubi, S., and White, M. F. Extensive lysine methylation in hyperthermophilic crenarchaea: potential implications for protein stability and recombinant enzymes. *Archaea* (2010).

[24] Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F., and Forterre, P. Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biol 6*, 5 (2005), R42.

[25] Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol. 6*, 3 (2008), 245–252.

[26] Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. *Sulfolobus*: a new genus of sulfur oxidising bacteria living at low pH and high temperature. *Arch. Microbiol. 84* (1972), 54–68.

[27] Brown, J. R., and Doolittle, W. F. Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev 61*, 4 (1997), 456–502.

[28] Brueckner, F., and Cramer, P. Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nat. Struct. Mol. Biol. 15*, 8 (2008), 811–818.

[29] Brun, I., Sentenac, A., and Werner, M. Dual role of the C34 subunit of RNA polymerase III in transcription initiation. *EMBO J. 16*, 18 (1997), 5730–5741.

[30] Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell 56*, 4 (1989), 549–561.

[31] Burke, T. W., and Kadonaga, J. T. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev 11*, 22 (1997), 3020–3031.

[32] Bushnell, D. A., Bamdad, C., and Kornberg, R. D. A minimal set of RNA polymerase II transcription protein interactions. *J. Biol. Chem. 271*, 33 (1996), 20170–20174.

[33] Bushnell, D. A., and Kornberg, R. D. Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc. Natl. Acad. Sci. 100*, 12 (2003), 6969–6973.

[34] Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoute, J., Brodsky, A. S., Keeton, E. K., Fertuck, K. C., Hall, G. F., Wang, Q., Bekiranov, S., Sementchenko, V., Fox, E. A., Silver, P. A., Gingeras, T. R., Liu, X. S., and Brown, M. Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet. 38*, 11 (2006), 1289–1297.

[35] Chasman, D. I., Flaherty, K. M., Sharp, P. A., and Kornberg, R. D. Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proc. Natl. Acad. Sci. 90*, 17 (1993), 8174–8178.

[36] Chen, C.-Y., Chang, C.-C., Yen, C.-F., Chiu, M. T.-K., and Chang, W.-H. Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc. Natl. Acad. Sci. 106*, 1 (2009), 127–132.

[37] Chen, H.-T., Warfield, L., and Hahn, S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nat. Struct. Mol. Biol. 14*, 8 (2007), 696–703.

[38] Chen, L., Brugger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H.-P.,

and Garrett, R. The genome of *Sulfolobus acidocaldarius*, a model organism of the crenarchaeota. *J. Bacteriol. 187* (2005), 4992–4999.

[39] Chen, Z. A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.-C., Nilges, M., Cramer, P., and Rappsilber, J. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J. 29*, 4 (2010), 717–726.

[40] Cheng, C., and Sharp, P. A. RNA polymerase II accumulation in the promoter-proximal region of the dihydrofolate reductase and gamma-actin genes. *Mol. Cell. Biol. 23*, 6 (2003), 1961–1967.

[41] Cheung, A. C. M., and Cramer, P. A movie of RNA polymerase II transcription. *Cell 149*, 7 (2012), 1431–1437.

[42] Cheung, A. C. M., Sainsbury, S., and Cramer, P. Structural basis of initial RNA polymerase II transcription. *EMBO J. 30*, 23 (2011), 4755–4763.

[43] Chung, W.-H., Craighead, J. L., Chang, W.-H., Ezeokonkwo, C., Bareket-Samish, A., Kornberg, R. D., and Asturias, F. J. RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. *Mol. Cell. 12*, 4 (2003), 1003–1013.

[44] Craighead, J. L., Chang, W.-h., and Asturias, F. J. Structure of yeast RNA polymerase II in solution: implications for enzyme regulation and interaction with promoter DNA. *Structure 10*, 8 (2002), 1117–1125.

[45] Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R., and Kornberg, R. D. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science 288*, 5466 (2000), 640–649.

[46] Cramer, P., Bushnell, D. A., and Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science 292*, 5523 (2001), 1863–1876.

[47] Darst, S. A., Edwards, A. M., Kubalek, E. W., and Kornberg, R. D. Three-dimensional structure of yeast RNA polymerase II at 16 A resolution. *Cell 66*, 1 (1991), 121–128.

[48] Darst, S. A., Kubalek, E. W., Edwards, A. M., and Kornberg, R. D. Two-dimensional and epitaxial crystallization of a mutant form of yeast RNA polymerase II. *J. Mol. Biol. 221*, 1 (1991), 347–357.

[49] Davis, S. E., Mooney, R. A., Kanin, E. I., Grass, J., Landick, R., and Ansari, A. Z. Mapping *E. coli* RNA polymerase and associated transcription factors and identifying promoters genome-wide. *Methods Enzymol. 498* (2011), 449–471.

[50] Demczuk, S., Harbers, M., and Vennstrom, B. Identification and analysis of all components of a gel retardation assay by combination with immunoblotting. *Proc. Natl. Acad. Sci. 90*, 7 (1993), 2574–2578.

[51] Duggin, I. G., McCallum, S. A., and Bell, S. D. Chromosome replication dynamics in the archaeon *Sulfolobus acidocaldarius*. *Proc. Natl. Acad. Sci. 105*, 43 (2008), 16737–16742.

[52] Edgell, D. R., and Doolittle, W. F. Archaea and the origin(s) of DNA replication proteins. *Cell 89*, 7 (1997), 995–998.

[53] Edwards, A. M., Darst, S. A., Feaver, W. J., Thompson, N. E., Burgess, R. R., and Kornberg, R. D. Purification and lipid-layer crystallization of yeast RNA polymerase II. *Proc. Natl. Acad. Sci. 87*, 6 (1990), 2122–2126.

[54] Edwards, A. M., Kane, C. M., Young, R. A., and Kornberg, R. D. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter *in vitro*. *J. Biol. Chem. 266*, 1 (1991), 71–75.

[55] Elkins, J. G., Podar, M., Graham, D. E., Makarova, K. S., Wolf, Y., Randau, L., Hedlund, B. P., Brochier-Armanet, C., Kunin, V., Anderson, I., Lapidus, A., Goltsman, E., Barry, K., Koonin, E. V., Hugenholtz, P., Kyrpides, N., Wanner, G., Richardson, P., Keller, M., and Stetter, K. O. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl. Acad. Sci. 105*, 23 (2008), 8102–8107.

[56] Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R. W. H., Sentenac, A., Muller, C. W., and Schoehn, G. Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. *Mol. Cell. 25*, 6 (2007), 813–823.

[57] Freire-Picos, M. A., Krishnamurthy, S., Sun, Z.-W., and Hampsey, M. Evidence that the Tfg1/Tfg2 dimer interface of TFIIF lies near the active center of the RNA polymerase II initiation complex. *Nucleic Acids Res. 33*, 16 (2005), 5045–5052.

[58] French, S. L., Santangelo, T. J., Beyer, A. L., and Reeve, J. N. Transcription and translation are coupled in Archaea. *Mol. Biol. Evol. 24*, 4 (2007), 893–895.

[59] Fu, J., Gnatt, A. L., Bushnell, D. A., Jensen, G. J., Thompson, N. E., Burgess, R. R., David, P. R., and Kornberg, R. D. Yeast RNA polymerase II at 5 A resolution. *Cell 98*, 6 (1999), 799–810.

[60] Geiduschek, E. P., and Kassavetis, G. A. The RNA polymerase III transcription apparatus. *J. Mol. Biol. 310*, 1 (2001), 1–26.

[61] Geiger, S. R., Lorenzen, K., Schreieck, A., Hanecker, P., Kostrewa, D., Heck, A. J. R., and Cramer, P. RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol. Cell. 39*, 4 (2010), 583–594.

[62] Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science 292*, 5523 (2001), 1876–1882.

[63] Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T., and Oshima, T. Evolution of the vacuolar h+-ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. 86*, 17 (1989), 6661–6665.

[64] Gribaldo, S., and Brochier-Armanet, C. The origin and evolution of Archaea: a state of the art. *Philos. Trans. R. Soc. Lond. B. Biol. Sci. 361*, 1470 (2006), 1007–1022.

[65] Gribaldo, S., and Cammarano, P. The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *J. Mol. Evol. 47*, 5 (1998), 508–516.

[66] Gribaldo, S., Poole, A. M., Daubin, V., Forterre, P., and Brochier-Armanet, C. The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat. Rev. Microbiol. 8*, 10 (2010), 743–752.

[67] Grogan, D. W. Exchange of genetic markers at extremely high temperatures in the archaeon *Sulfolobus acidocaldarius*. *J. Bacteriol. 178*, 3207-3211 (1996).

[68] Grohmann, D., Hirtreiter, A., and Werner, F. RNAP subunits F/E (RPB4/7) are stably associated with archaeal RNA polymerase: using fluorescence anisotropy to monitor RNAP assembly *in vitro*. *Biochem J. 421*, 3 (2009), 339–343.

[69] Grove, A., Galeone, A., Yu, E., Mayol, L., and Geiduschek, E. P. Affinity, stability and polarity of binding of the TATA binding protein governed by flexure at the TATA box. *J. Mol. Biol. Mol Biol 282*, 4 (1998), 731–739.

[70] Guo, L., Feng, Y., Zhang, Z., Yao, H., Luo, Y., Wang, J., and Huang, L. Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among crenarchaea. *Nucleic Acids Res. 36*, 4 (2008), 1129–1137.

[71] Ha, I., Roberts, S., Maldonado, E., Sun, X., Kim, L. U., Green, M., and Reinberg, D. Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II. *Genes Dev. 7*, 6 (1993), 1021–1032.

[72] Hanzelka, B. L., Darcy, T. J., and Reeve, J. N. TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIEalpha. *J. Bacteriol. 183*, 5 (2001), 1813–1818.

[73] Hausner, W., Wettach, J., Hethke, C., and Thomm, M. Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. *J. Biol. Chem. 271*, 47 (1996), 30144–30148.

[74] He, Y., Fang, J., Taatjes, D. J., and Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature 495*, 7442, 481–486.

[75] Heidemann, M., Hintermair, C., Voss, K., and Eick, D. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim. Biophys. Acta* (2012).

[76] Herring, C. D., Raffaelle, M., Allen, T. E., Kanin, E. I., Landick, R., Ansari, A. Z., and Palsson, B. O. Immobilization of

E. coli RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *J. Bacteriol. 187*, 17 (2005), 6166–6174.

[77] Hethke, C., Geerling, A. C., Hausner, W., de Vos, W. M., and Thomm, M. A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. *Nucleic Acids Res. 24*, 12 (1996), 2369–2376.

[78] Hieb, A. R., Baran, S., Goodrich, J. A., and Kugel, J. F. An 8 nt RNA triggers a rate-limiting shift of RNA polymerase II complexes into elongation. *EMBO J. 25*, 13 (2006), 3100–3109.

[79] Hirata, A., Klein, B. J., and Murakami, K. S. The X-ray crystal structure of RNA polymerase from Archaea. *Nature 451*, 7180 (2008), 851–854.

[80] Holstege, F. C., van der Vliet, P. C., and Timmers, H. T. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J. 15*, 7 (1996), 1666–1677.

[81] Hsin, J.-P., and Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev. 26*, 19 (2012), 2119–2137.

[82] Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., and Stetter, K. O. A new phylum of archaea represented by a nanosized hyperthermophilic symbiont. *Nature 417*, 6884 (2002), 63–67.

[83] Hudepohl, U., Reiter, W. D., and Zillig, W. *In vitro* transcription of two rRNA genes of the archaebacterium *Sulfolobus* sp. B12 indicates a factor requirement for specific initiation. *Proc. Natl. Acad. Sci 87*, 15 (1990), 5851–5855.

[84] Ishikawa, S., Oshima, T., Kurokawa, K., Kusuya, Y., and Ogasawara, N. RNA polymerase trafficking in *Bacillus subtilis cells*. *J. Bacteriol. 192*, 21 (2010), 5778–5787.

[85] Jasiak, A. J., Armache, K.-J., Martens, B., Jansen, R.-P., and Cramer, P. Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol. Cell. 23*, 1 (2006), 71–81.

[86] Jasiak, A. J., Hartmann, H., Karakasili, E., Kalocsay, M., Flatley, A., Kremmer, E., Strasser, K., Martin, D. E., Soding, J., and Cramer, P. Genome-associated rna polymerase ii includes the dissociable rpb4/7 subcomplex. *J Biol Chem 283*, 39, 26423–26427.

[87] Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J., and Dickerson, R. E. How proteins recognize the TATA box. *J. Mol. Biol. 261*, 2 (1996), 239–254.

[88] Kaplan, C. D., Larsson, K.-M., and Kornberg, R. D. The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol. Cell. 30*, 5 (2008), 547–556.

[89] Kassavetis, G. A., Joazeiro, C. A., Pisano, M., Geiduschek, E. P., Colbert, T., Hahn, S., and Blanco, J. A. The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB. *Cell 71*, 6 (1992), 1055–1064.

[90] Kettenberger, H., Armache, K.-J., and Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell. 16*, 6 (2004), 955–965.

[91] Khoo, B., Brophy, B., and Jackson, S. P. Conserved functional domains of the RNA polymerase III general transcription factor BRF. *Genes. Dev. 8*, 23 (1994), 2879–2890.

[92] Kim, J. L., Nikolov, D. B., and Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature 365*, 6446 (1993), 520–527.

[93] Kim, T. K., Ebright, R. H., and Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science 288*, 5470 (2000), 1418–1422.

[94] Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature 365*, 6446 (1993), 512–520.

[95] Koleske, A. J., and Young, R. A. An RNA polymerase II holoenzyme responsive to activators. *Nature 368*, 6470 (1994), 466–469.

[96] Korkhin, Y., Unligil, U. M., Littlefield, O., Nelson, P. J., Stuart, D. I., Sigler, P. B., Bell, S. D., and Abrescia, N. G. A. Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol 7*, 5 (2009), e1000102.

[97] Kosa, P. F., Ghosh, G., DeDecker, B. S., and Sigler, P. B. The 2.1-A crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc. Natl. Acad. Sci. 94*, 12 (1997), 6042–6047.

[98] Kostrewa, D., Zeller, M. E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature 462*, 7271 (2009), 323–330.

[99] Krumm, A., Hickey, L. B., and Groudine, M. Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. *Genes Dev. 9*, 5 (1995), 559–572.

[100] Kuhn, C.-D., Geiger, S. R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. Functional architecture of RNA polymerase I. *Cell 131*, 7 (2007), 1260–1272.

[101] Kusser, A. G., Bertero, M. G., Naji, S., Becker, T., Thomm, M., Beckmann, R., and Cramer, P. Structure of an archaeal RNA polymerase. *J. Mol. Biol. 376*, 2 (2008), 303–307.

[102] Kwapisz, M., Beckouet, F., and Thuriaux, P. Early evolution of eukaryotic DNA-dependent RNA polymerases. *Trends Genet. 24*, 5 (2008), 211–215.

[103] Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., and Ebright, R. H. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev. 12*, 1 (1998), 34–44.

[104] Lanzendorfer, M., Langer, D., Hain, J., Klenk, H., Holz, I., Arnoldammer, I., and Zillig, W. Structure and function of the DNA-dependent RNA-polymerase of *Sulfolobus*. *Systematic and Applied Microbiol. 16*, 4 (1994), 656–664.

[105] Lecompte, O., Ripp, R., Thierry, J.-C., Moras, D., and Poch, O. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res. 30*, 24 (2002), 5382–5390.

[106] Leipe, D. D., Aravind, L., and Koonin, E. V. Did DNA replication evolve twice independently? *Nucleic Acids Res. 27*, 17 (1999), 3389–3401.

[107] Lessner, F. H., Jennings, M. E., Hirata, A., Duin, E. C., and Lessner, D. J. Subunit d of rna polymerase from methanosarcina acetivorans contains two oxygen-labile [4fe-4s] clusters: implications for oxidant-dependent regulation of transcription. *J Biol Chem 287*, 22 (2012 May 25), 18510–18523.

[108] Leuther, K. K., Bushnell, D. A., and Kornberg, R. D. Two-dimensional crystallography of TFIIB- and IIE-RNA polymerase II complexes:

implications for start site selection and initiation complex formation. *Cell 85*, 5 (1996), 773–779.

[109] Lindas, A.-C., Karlsson, E. A., Lindgren, M. T., Ettema, T. J. G., and Bernander, R. A unique cell division machinery in the archaea. *Proc. Natl. Acad. Sci. 105*, 48 (2008), 18942–18946.

[110] Littlefield, O., Korkhin, Y., and Sigler, P. B. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc. Natl. Acad. Sci. 96*, 24 (1999), 13668–13673.

[111] Liu, C. L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S. L., Friedman, N., and Rando, O. J. Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol 3*, 10 (2005), e328.

[112] Liu, X., Bushnell, D. A., Silva, D.-A., Huang, X., and Kornberg, R. D. Initiation complex structure and promoter proofreading. *Science 333*, 6042 (2011), 633–637.

[113] Liu, X., Bushnell, D. A., Wang, D., Calero, G., and Kornberg, R. D. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science 327*, 5962 (2010), 206–209.

[114] Londei, P. Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol. Rev. 29*, 2 (2005), 185–200.

[115] Lorenzen, K., Vannini, A., Cramer, P., and Heck, A. J. R. Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure 15*, 10 (2007), 1237–1245.

[116] Magill, C. P., Jackson, S. P., and Bell, S. D. Identification of a conserved archaeal RNA polymerase subunit contacted by the basal transcription factor TFB. *J. Biol. Chem. 276*, 50 (2001), 46693–46696.

[117] Mai, V. Q., Chen, X., Hong, R., and Huang, L. Small abundant DNA binding proteins from the thermoacidophilic archaeon *Sulfolobus shibatae* constrain negative DNA supercoils. *J. Bacteriol. 180*, 9 (1998), 2560–2563.

[118] Marinsek, N., Barry, E. R., Makarova, K. S., Dionne, I., Koonin, E. V., and Bell, S. D. GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep. 7*, 5 (2006), 539–545.

[119] Meinhart, A., Blobel, J., and Cramer, P. An extended winged helix domain in general transcription factor E/IIE alpha. *J. Biol. Chem. 278*, 48 (2003), 48267–48274.

[120] Meinhart, A., and Cramer, P. Recognition of rna polymerase ii carboxy-terminal domain by 3'-rna-processing factors. *Nature 430*, 6996 (2004), 223–226.

[121] Meka, H., Werner, F., Cordell, S. C., Onesti, S., and Brick, P. Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. *Nucleic Acids Res. 33*, 19 (2005), 6435–6444.

[122] Micorescu, M., Grunberg, S., Franke, A., Cramer, P., Thomm, M., and Bartlett, M. Archaeal transcription: function of an alternative transcription factor B from *Pyrococcus furiosus*. *J. Bacteriol. 190*, 1 (2008), 157–167.

[123] Miller, G., and Hahn, S. A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nat. Struct. Mol. Biol. 13*, 7 (2006), 603–610.

[124] Mooney, R. A., Davis, S. E., Peters, J. M., Rowland, J. L., Ansari, A. Z., and Landick, R. Regulator trafficking on bacterial transcription units *in vivo*. *Mol. Cell. 33*, 1 (2009), 97–108.

[125] Nacheva, G. A., and Berzal-Herranz, A. Preventing nondesired RNA-primed RNA extension catalyzed by T7 RNA polymerase. *Eur. J. Biochem. 270*, 7 (2003), 1458–1465.

[126] Naji, S., Bertero, M. G., Spitalny, P., Cramer, P., and Thomm, M. Structure-function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement. *Nucleic Acids Res. 36*, 2 (2008), 676–687.

[127] Naji, S., Grunberg, S., and Thomm, M. The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J. Biol. Chem. 282*, 15 (2007), 11047–11057.

[128] Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature 377*, 6545 (1995), 119–128.

[129] Nikolov, D. B., Hu, S. H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N. H., Roeder, R. G., and Burley, S. K. Crystal structure of TFIID TATA-box binding protein. *Nature 360*, 6399 (1992), 40–46.

[130] Orlicky, S. M., Tran, P. T., Sayre, M. H., and Edwards, A. M. Dissociable Rpb4-Rpb7 subassembly of RNA polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. *J. Biol. Chem. 276*, 13 (2001), 10097–10102.

[131] Ossipow, V., Tassan, J. P., Nigg, E. A., and Schibler, U. A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell 83*, 1 (1995), 137–146.

[132] Ouhammouch, M., and Geiduschek, E. P. An expanding family of archaeal transcriptional activators. *Proc. Natl. Acad. Sci. 102*, 43 (2005), 15423–15428.

[133] Ouhammouch, M., Werner, F., Weinzierl, R. O. J., and Geiduschek, E. P. A fully recombinant system for activator-dependent archaeal transcription. *J. Biol. Chem. 279*, 50 (2004), 51719–51721.

[134] Peyroche, G., Levillain, E., Siaut, M., Callebaut, I., Schultz, P., Sentenac, A., Riva, M., and Carles, C. The A14-A43 heterodimer subunit in yeast RNA pol I and their relationship to Rpb4-Rpb7 pol II subunits. *Proc. Natl. Acad. Sci. 99*, 23 (2002), 14670–14675.

[135] Poglitsch, C. L., Meredith, G. D., Gnatt, A. L., Jensen, G. J., Chang, W. H., Fu, J., and Kornberg, R. D. Electron crystal structure of an RNA polymerase II transcription elongation complex. *Cell 98*, 6 (1999), 791–798.

[136] Proshkina, G. M., Shematorova, E. K., Proshkin, S. A., Zaros, C., Thuriaux, P., and Shpakovski, G. V. Ancient origin, functional conservation and fast evolution of DNA-dependent RNA polymerase III. *Nucleic Acids Res. 34*, 13 (2006), 3615–3624.

[137] Qureshi, S. A., Baumann, P., Rowlands, T., Khoo, B., and Jackson, S. P. Cloning and functional analysis of the TATA binding protein from *Sulfolobus shibatae*. *Nucleic Acids Res. 23*, 10 (1995), 1775–1781.

[138] Qureshi, S. A., Bell, S. D., and Jackson, S. P. Factor requirements for transcription in the archaeon *Sulfolobus shibatae*. *EMBO J. 16*, 10 (1997), 2927–2936.

[139] Qureshi, S. A., and Jackson, S. P. Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol. Cell. 1*, 3 (1998), 389–400.

[140] Radonjic, M., Andrau, J.-C., Lijnzaad, P., Kemmeren, P., Kockelkorn, T. T. J. P., van Leenen, D., van Berkum, N. L., and Holstege, F. C. P. Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon *S. cerevisiae* stationary phase exit. *Mol. Cell. 18*, 2 (2005), 171–183.

[141] Reppas, N. B., Wade, J. T., Church, G. M., and Struhl, K. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Mol. Cell. 24*, 5 (2006), 747–757.

[142] Richards, J. D., Cubeddu, L., Roberts, J., Liu, H., and White, M. F. The archaeal XPB protein is a ssDNA-dependent ATPase with a novel partner. *J. Mol. Biol. 376*, 3 (2008), 634–644.

[143] Roberts, D. N., Stewart, A. J., Huff, J. T., and Cairns, B. R. The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. *Proc. Natl. Acad. Sci. 100*, 25 (2003), 14695–14700.

[144] Rodriguez-Wells, V., Plotch, S. J., and DeStefano, J. J. Primer-dependent synthesis by poliovirus RNA-dependent RNA polymerase (3D(pol)). *Nucleic Acids Res. 29*, 13 (2001), 2715–2724.

[145] Rougvie, A. E., and Lis, J. T. The RNA polymerase II molecule at the 5' end of the uninduced *hsp70* gene of *D. melanogaster* is transcriptionally engaged. *Cell 54*, 6 (1988), 795–804.

[146] Rowlands, T., Baumann, P., and Jackson, S. P. The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. *Science 264*, 5163 (1994), 1326–1329.

[147] Rudolf, J., Rouillon, C., Schwarz-Linek, U., and White, M. F. The helicase XPD unwinds bubble structures and is not stalled by DNA lesions

removed by the nucleotide excision repair pathway. *Nucleic Acids Res. 38*, 3 (2010), 931–941.

[148] Rusakova, E. E., Tunitskaya, V. L., Memelova, L. V., Kochetkova, S. V., Kostyuk, D. A., and Kochetkov, S. N. Mutant T7 RNA polymerase is capable of catalyzing DNA primer extension reaction. *FEBS Lett. 423*, 2 (1998), 189–192.

[149] Samson, R. Y., and Bell, S. D. Ancient ESCRTs and the evolution of binary fission. *Trends Microbiol 17*, 11 (2009 Nov), 507–513.

[150] Samson, R. Y., Obita, T., Freund, S. M., Williams, R. L., and Bell, S. D. A role for the ESCRT system in cell division in archaea. *Science 322*, 5908 (2008), 1710–1713.

[151] Sanso, M., and Fisher, R. P. Pause, play, repeat: Cdks push rnap ii's buttons. *Transcription 4*, 4 (2013 Jun 11).

[152] Starr, D. B., Hoopes, B. C., and Hawley, D. K. DNA bending is an important component of site-specific recognition by the TATA binding protein. *J. Mol. Biol. 250*, 4 (1995), 434–446.

[153] Steinmetz, E. J., Warren, C. L., Kuehner, J. N., Panbehi, B., Ansari, A. Z., and Brow, D. A. Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol. Cell. 24*, 5 (2006), 735–746.

[154] Steitz, T. A. A mechanism for all polymerases. *Nature 391*, 6664 (1998), 231–232.

[155] Takiya, S., Saito, S., Yokoyama, T., Matsumoto, D., Aizawa, T., Kamiya, M., Demura, M., and Kawano, K. DNA-binding property of the novel DNA-binding domain STPR in FMBP-1 of the silkworm *Bombyx mori*. *J. Biochem. 146*, 1 (2009), 103–111.

[156] Tan, S., Hunziker, Y., Sargent, D. F., and Richmond, T. J. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature 381*, 6578 (1996), 127–151.

[157] te Velthuis, A. J. W., van den Worm, S. H. E., and Snijder, E. J. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both *de novo* initiation and primer extension. *Nucleic Acids Res. 40*, 4 (2012), 1737–1747.

[158] Todone, F., Brick, P., Werner, F., Weinzierl, R. O., and Onesti, S. Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol. Cell. 8*, 5 (2001), 1137–1143.

[159] Ujvari, A., and Luse, D. S. RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat. Struct. Mol. Biol. 13*, 1 (2006), 49–54.

[160] Vannini, A., and Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell. 45*, 4 (2012), 439–446.

[161] Vannini, A., Ringel, R., Kusser, A. G., Berninghausen, O., Kassavetis, G. A., and Cramer, P. Molecular basis of RNA polymerase III transcription repression by Maf1. *Cell 143*, 1 (2010), 59–70.

[162] Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H., and Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature 448*, 7150 (2007), 157–162.

[163] Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. Structural basis for substrate loading in bacterial RNA polymerase. *Nature 448*, 7150 (2007), 163–168.

[164] Verrijzer, C. P., Chen, J. L., Yokomori, K., and Tjian, R. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell 81*, 7 (1995), 1115–1125.

[165] Volff, J. N., and Altenbuchner, J. A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. *FEMS Microbiol Lett 186*, 2 (2000), 143–150.

[166] Wade, J. T., and Struhl, K. Association of RNA polymerase with transcribed regions in *E. coli*. *Proc. Natl. Acad. Sci. 101*, 51 (2004), 17777–17782.

[167] Wade, J. T., and Struhl, K. The transition from transcriptional initiation to elongation. *Curr. Opin. Genet. Dev. 18*, 2 (2008), 130–136.

[168] Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D., and Kornberg, R. D. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell 127*, 5 (2006), 941–954.

[169] Wang, Z., and Roeder, R. G. Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. *Genes Dev. 11*, 10 (1997), 1315–1326.

[170] Werner, F. Structure and function of archaeal RNA polymerases. *Mol. Microbiol. 65*, 6 (2007), 1395–1404.

[171] Werner, F. Structural evolution of multisubunit RNA polymerases. *Trends Microbiol. 16*, 6 (2008), 247–250.

[172] Werner, F., and Weinzierl, R. O. J. Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol. Cell. Biol. 25*, 18 (2005), 8344–8355.

[173] Werner, M., Chaussivert, N., Willis, I. M., and Sentenac, A. Interaction between a complex of RNA polymerase III subunits and the 70-

kda component of transcription factor IIIB. *J Biol Chem 268*, 28 (1993), 20721–20724.

[174] Westover, K. D., Bushnell, D. A., and Kornberg, R. D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell 119*, 4 (2004), 481–489.

[175] Westover, K. D., Bushnell, D. A., and Kornberg, R. D. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science 303*, 5660 (2004), 1014–1016.

[176] Woese, C., and Fox, G. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. 74*, 11 (1977), 5088–5090.

[177] Woese, C. R. Bacterial evolution. *Microbiological Reviews 51*, 2 (1987), 221–271.

[178] Woese, C. R., Kandler, O., and Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. 87* (1990), 4576–4579.

[179] Wojtas, M., Peralta, B., Ondiviela, M., Mogni, M., Bell, S. D., and Abrescia, N. G. A. Archaeal RNA polymerase: the influence of the protruding stalk in crystal packing and preliminary biophysical analysis of the Rpo13 subunit. *Biochem. Soc. Trans. 39*, 1 (2011), 25–30.

[180] Wojtas, M. N., Mogni, M., Millet, O., Bell, S. D., and Abrescia, N. G. A. Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA. *Nucleic Acids Res.* (2012).

[181] Yan, Q., Moreland, R. J., Conaway, J. W., and Conaway, R. C. Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J. Biol. Chem. 274*, 50 (1999), 35668–35675.

[182] Zaros, C., Briand, J.-F., Boulard, Y., Labarre-Mariotte, S., Garcia-Lopez, M. C., Thuriaux, P., and Navarro, F. Functional organization of the Rpb5 subunit shared by the three yeast RNA polymerases. *Nucleic Acids Res. 35*, 2 (2007), 634–647.

[183] Ziegler, L. M., Khaperskyy, D. A., Ammerman, M. L., and Ponticelli, A. S. Yeast RNA polymerase II lacking the Rpb9 subunit is impaired for interaction with transcription factor IIF. *J. Biol. Chem. 278*, 49 (2003), 48950–48956.

[184] Zillig, W. Comparative biochemistry of archaea and bacteria. *Current Opinion in Genetics and Development 1* (1991), 544–551.